AD_____

Award Number: W81XWH-10-1-1007

TITLE: Molecular Profiles for Lung Cancer Pathogenesis and Detection in US Veterans

PRINCIPAL INVESTIGATOR: Ignacio Wistuba, M.D.

CONTRACTING ORGANIZATION: MD Anderson Cancer Center
Houston, TX 77030-4009

REPORT DATE: October 2011

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
October 2011	Annual	20September 2010 – 19 September 2011

4. TITLE AND SUBTITLE

Molecular Profiles for Lung Cancer Pathogenesis and Detection in US Veterans

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-10-1-1007

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Ignacio Wistuba, M.D.

E-Mail: iiwistuba@mdanderson.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

MD Anderson Cancer Center
Houston, TX 77030-4009

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

.
Different localized field effect phenomenon occurs in patients with lung adenocarcinoma and squamous cell carcinoma histologies.

15. SUBJECT TERMS
Field cancerization, lung cancer

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	62	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

Molecular Profiles for Lung Cancer Pathogenesis and Detection in U.S. Veterans.

INTRODUCTION:

Lung cancer continues to be the leading cause of cancer-related death in both men and women in the United States. The majority of lung cancers are non-small cell lung cancers (NSCLCs) that include squamous cell carcinomas (SCCs) and adenocarcinomas. Lung cancer mortality is high in part because most cancers are diagnosed after regional or distant spread of the disease had already occurred and due to the lack of reliable biomarkers for early detection and risk assessment. The identification of new effective early biomarkers will improve clinical management of lung cancer and is linked to better understanding of the molecular events associated with the development and progression of the disease.

It has been suggested that histologically normal-appearing tissue adjacent to neoplastic lesions display molecular abnormalities some of which are in common with those in the tumors. This phenomenon, termed field of cancerization, was later shown to be evident in various epithelial cell malignancies, including lung cancer. Loss of heterozygosity (LOH) events are frequent in cells obtained from bronchial brushings of normal and abnormal lungs from patients undergoing diagnostic bronchoscopy and were detected in cells from the ipsilateral and contralateral lungs. More recently, global mRNA and microRNA (miRNA) expression profiles have been described in the normal-appearing bronchial epithelium of health smokers. In addition, modulation of global gene expression in the normal epithelium in health smokers is similar in the large and small airways and the smoking-induced alterations are mirrored in the epithelia of the mainstem bronchus, buccal and nasal cavities.

In this program, in Specific Aim 1, high-throughput microarray mRNA and miRNA expression analyses will be performed on cytological specimens (brushings) obtained at intraoperative bronchoscopy from the main carina and main ipsilateral bronchus, as well as on specimens obtained at lobectomy procedures from the main lobe bronchus (adjacent to SCCs), sub-segmental bronchus (adjacent to adenocarcinomas) and from the resected NSCLC tumors. We will compare and contrast global gene expression, both mRNA and miRNA, patterns across all the specimens from the entire field and corresponding NSCLC tumors. We seek to derive lung adenocarcinoma and SCC field cancerization signatures signifying the differential mRNA and miRNA expression patterns between the carina and the subsegmental bronchus and main lobe bronchus, respectively. In addition, similar expression profiles between the carina and resected NSCLC tumors will be integrated with available gene expression data of bronchial brushings from the main carina isolated at various time points post-surgery from 40 NSCLC patients; Department of Defense (DoD) VITAL patients. Lastly, functional pathway analysis will be performed to organize differentially expressed genes into topological biological networks in association with miRNA expression. Promising markers derived from this innovative study will be validated at the mRNA and protein level in histological tissue specimens and may be tested in future studies to determine their potential role in improving risk assessment, providing new targets for novel chemopreventive agents and selecting NSCLC patients who may benefit from chemopreventive interventions to prevent disease recurrence.

In Specific Aim 2 we are evaluating the role of airway epithelium tumor-initiating stem/progenitor cells in current and former smokers. The purpose of this Aim is to profile the airways of patients with NSCLC to detect a subpopulation of tumor-initiating cells that will lead to the identification of candidate biomarkers. These biomarkers are important for understanding early events of lung

cancer pathogenesis, relevant for identifying persons at highest risk of developing lung cancer, and useful in early detection of lung cancer. To achieve this Aim we are developing novel methods to detect tumor-initiating cells in NSCLC samples and performing high throughput sequencing to determine gene expression profiles that inform lung carcinogenesis and develop biomarkers of lung cancer. In Specific Aim 3 we will use gene signatures developed in Specific Aims 1 and 2 to test airway-based mRNA and microRNA biomarkers capable of diagnosing lung cancer in current and former smokers in minimally invasive sites.

This report details the comprehensive annual progress of the consortium made during the first year of research.

PROGRESS REPORT:

Specific Aim 1: **To increase our understanding of the molecular basis of the pathogenesis of lung cancer in the "field cancerization" that develops in current and former smokers.**

Summary of Research Findings

A. Gene and protein expression analysis of bronchial epithelial samples obtained from bronchoscopy from NSCLC patients (Sub-specific Aim 1A, 1C and 1D): This analysis was performed on samples obtained from patients enrolled to the Vanguard clinical trial supported by the DoD VITAL grant (W81XWH-04-1-0412, PI Dr. W.K. Hong), and the gene profiling work and analysis, and the protein validation study, was partially supported by the Lung Cancer Consortium grant reported here.

Gene Expression Analysis: Patients on the prospective Vanguard study had definitively treated ES (I/II) NSCLC and were current or former smokers. Patients had bronchoscopies with brushings obtained from the main carina (MC), airways adjacent (ADJ) to the previously resected tumor and from airways distant from the tumor in the ipsilateral (NON-ADJ) and contralateral (CONTRA) lung at baseline, 12, 24 and 36 months following resective surgery (**Figure 1**). Nineteen patients were selected for the study based on airway sampling of at least five different regions per time point and continuously up to 24 or 36 months (391 airway samples from nineteen patients). Total RNA was isolated from brushings using the RNeasy Mini kit (Qiagen) according to the manufacturer's instructions. Due to the paucity of the material, we employed the Nugen WT-Ovation system for RNA amplification (Nugen Technologies, San Carlos, CA). Synthesis of single-stranded DNA, fragmentation and biotin labeling was performed using the WT-Ovation Pico RNA amplification system, WT-

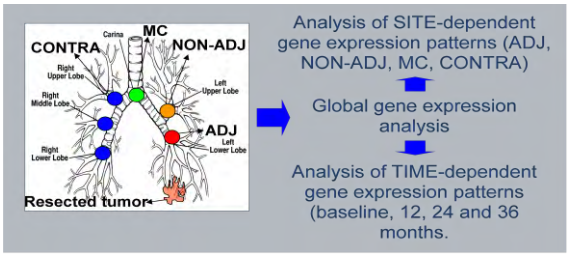

Figure 1. Schematic illustrating sampling and collection of bronchial brushes from different airway sites at different time points following resective surgery from early stage smoker NSCLC patients enrolled in the phase II cancer surveillance Vanguard study.

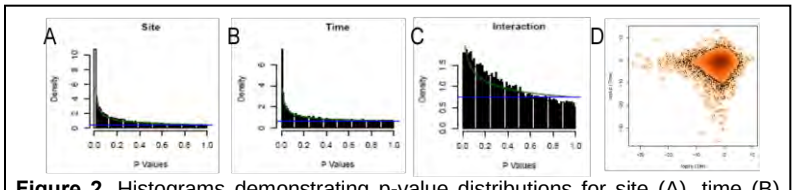

Figure 2. Histograms demonstrating p-value distributions for site (A), time (B) and interaction between both after fitting beta-uniform-mixture (BUM) models (3). D. Smoothed scatter plot of transformed p-values from site and time.

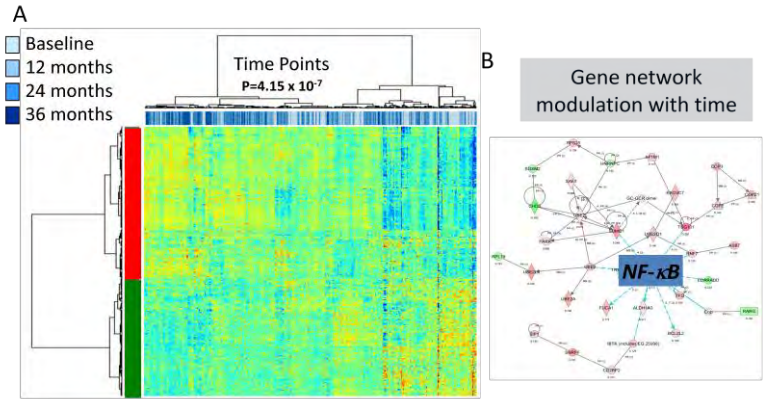

Figure 3. Gene expression profiles are different in a time-dependent manner in the field of cancerization. Hierarchical clustering (A) and functional gene-interaction network analysis (B) of genes (n=1395) differentially expressed continuously with time.

Ovation Exon-Module and Encore™ Biotin Module (NuGEN), respectively, according to the manufacturer's instructions. 2-2.5 micrograms of biotin-labeled DNA was then hybridized to Affymetrix Human Gene 1.0 ST arrays from Affymetrix. Analysis began by construction of a mixed-effects model that incorporated information on the site and time (continuous) the bronchial brushing was obtained as fixed effects. Characteristics of fixed effects and their interaction, in terms of number of genes differentially affected by the effect, were examined by fitting beta-uniform-models (BUMs) based on p-value distributions of all genes according to the fixed-effect (**Figure 2**). Histograms of p-value distribution of fixed effects (**Figures 2A, 2B and 2C**) and a smoothened scatter plot of transformed p-values of both site and time fixed effects (**Figure 2D**) demonstrate that there is little evidence for any interaction between site and time. These findings suggest that, although the airway brushings were collected at different sites and at time points following respective surgery from each patient, all airway samples (n=391) can be utilized independently to unravel genes differentially expressed by site from original resected tumor and by time following removal of the same tumor in all patients. Subsequently, time and site-dependent field of cancerization differential expression profiles were identified based on a false discovery rate (FDR) cut-off of 5% and 1% based on p-value distributions, respectively, studied by hierarchical clustering and functionally analyzed using network analysis. 1,395 genes were determined to be differentially expressed with time in the cancerization fields.

Hierarchical clustering analysis using these genes demonstrated that samples (n=391) were divided into two clusters or branches which were significantly unbalanced with respect to time with the majority of the baseline and 36 months samples separated ($p<0.001$ of the Fisher's Exact test) (**Figure 3A**). Moreover, functional analysis of these genes showed that a nuclear factor-κB (NF-κB)-mediated gene-network was most significantly elevated ($p<0.001$) with time (**Figure 3B**).

1,165 gene features were differentially expressed by site. Two-dimensional clustering of these genes and samples showed distinct

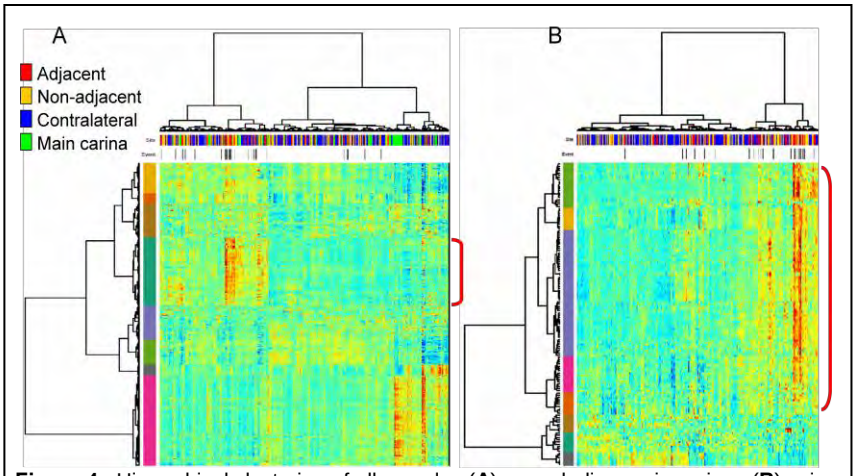

Figure 4. Hierarchical clustering of all samples (**A**) or excluding main carinas (**B**) using genes differentially expressed by site. Recurrences (black) and suspicions (grey) are labeled based on adjacent airways. Genes with highest expression in adjacent airways are indicated by the red bars.

classes of differential expression and revealed two main sample clusters with significant separation of ADJ samples from MCs and non-adjacent CONTRA airway samples (p=0.003) (**Figure 4A**). Similar results were obtained when the main carina samples were excluded from analysis of genes differentially expressed by site in relation to the original resected tumor (**Figure 4B**). Using both site-dependent analyses, genes were identified that exhibited highest average expression in airways adjacent to tumors (cluster of genes highlighted by red bars) (**Figures 4A and 4B**). It is worthwhile to mention that, following two-dimensional clustering of the site-dependent differentially expressed genes and airway samples, adjacent airways isolated from patients with recurrence or suspicion of recurrence, (black, recurrence; grey, suspicion of recurrence), exhibited on average elevated expression of the highest-in-adjacent airway gene cluster compared to adjacent airways isolated from patients with no events in recurrence (**Figures 4A and 4B**). These findings suggest that differential gene expression patterns, by site from original tumor, in the field of cancerization of early stage patients may be associated with recurrence or second primary tumor development.

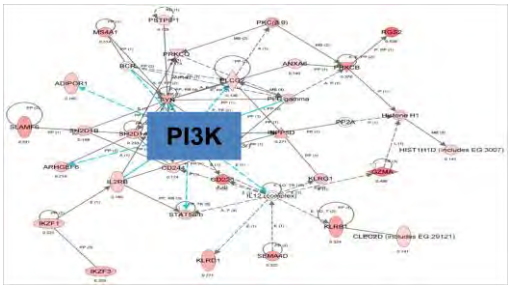

Figure 5. Network analysis of cluster of genes depicted in **Figure 4** with highest average expression in airways adjacent to original resected tumors. PI3K-mediated network is shown as most significantly up-regulated in adjacent airways . red, up-regulated in adjacent airways relative to non-adjacent or contralateral airways.

To further understand the relevance of this gene cluster with highest expression in adjacent airways and isolated from patients with recurrence, we performed functional pathway analysis of the genes using the knowledge-dependent analysis, Ingenuity Pathways analysis (IPA). Functional analysis of the highest-in-adjacent airway genes revealed that gene-networks mediated by PI3K, NF-κB, and ERK1/2 had significantly elevated ($p<0.001$) function in ADJ airway samples, with a gene-interaction network mediated by PI3K being most significantly elevated in function, as predicted by the IPA software, based on number of genes differentially expressed within the network and topological interactions among the same genes (**Figure 5**). These findings suggest that the aforementioned canonical cell signaling pathways, in particular the PI3K survival pathway may be highly relevant biologically to the molecular pathogenesis of NSCLC, and clinically to predict recurrence or second primary tumor development in early stage patients definitively cured by resective surgery.

We further identified genes differentially expressed by site using different statistical methods. We identified site-dependent gene expression patterns using paired t-test analysis of the 19 NSCLC patients based on differences in expression between adjacent and non-adjacent/contralateral airways. An average ADJ expression score and average non-adjacent/contralateral (NON-ADJ) score was measured for each gene based on all available airway samples per patient. Hierarchical clustering (**Figure 6A**) and principal component (**Figure 6B**) analysis of patients (n=19) based on genes with significant expression

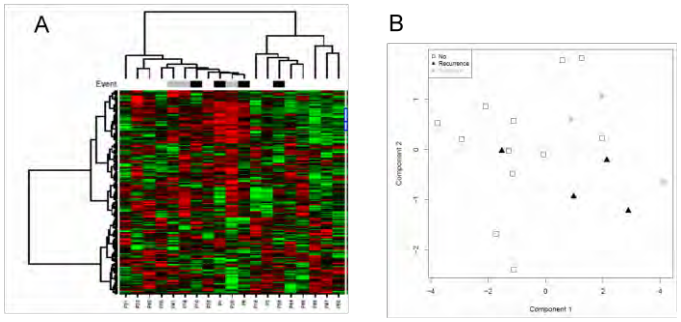

Figure 6. Field of cancerization molecular profiles are associated with patterns of recurrence. Genes with significant expression differences between adjacent and non-adjacent airways (ADJ/NON-ADJ) were used to analyze the 19 pts by clustering (A) and principal component analysis (B). Black: recurrence; grey: suspicion of recurrence.

differences (by paired t-tests) between ADJ and NON-ADJ samples revealed two main clusters with three of four relapses located in one sub-cluster suggesting potential associations between field of cancerization expression profiles and lung cancer relapse.

Protein Expression Analysis: Our findings on the significant modulation of a PI3K-mediated gene-interaction network in adjacent airways compared to other airway brushings (**Figure 5**) prompted us to examine the immunohistochemical (IHC) expression of phosphoAKT (Threonine 308) in available airway biopsy specimens corresponding to the bronchial brushings; we previously analyzed the transcriptome by microarray profiling. AKT is phosphorylated at two major sites, serine 473 and threonine 308, the latter site being phosphorylated through PDK1 following PI3K activation, thus acting as a key surrogate for this pathway activation. We assessed the expression of phosphoAKT-Thr308 by IHC analysis in available and eligible airway biopsy specimens (n=324). Antigen retrieval was performed using the Dako target retrieval system at a PH of 6. Intrinsic peroxidase activity was blocked by 3%methanol and hydrogen peroxide for 15 min and serum-free protein block (Dako) was used for 7 min for blocking non-specific antibody binding. Slides were then incubated at room temperature for 90 min with the primary antibody raised against pAKT-Thr308 (dilution 1:200, clone C31E5E, catalog number 2965, Cell signaling Technology, Danvers, MA). After three washes in Tris-buffered saline, slides were incubated for 30 min with Dako Envision+ Dual Link at room temperature. Following three additional washes, slides were incubated with Dako chromogen substrate for 5 min and were counterstained with hemotoxylin for another 5 min. The intensity and extent of cytoplasmic and nuclear pAKT-Thr308 immunostaining was evaluated using a light microscope (magnification, x20). Cytoplasmic expression was quantified using a four-value intensity score (0, none; 1, weak; 2, moderate and 3, strong) and the percentage (0-100%) of the extent of reactivity). A final cytoplasmic expression score for pAKT-Thr308 was obtained by multiplying the intensity and reactivity extension values (range, 0-300). Nuclear expression was quantified using the percentage of extent of reactivity, which gave rise to a nuclear expression score for pAKT-Thr308 (range, 0-100). Representative pAKT-Thr308 immunostaining (upper, strong; bottom, weak) is depicted in **Figure 7A**.

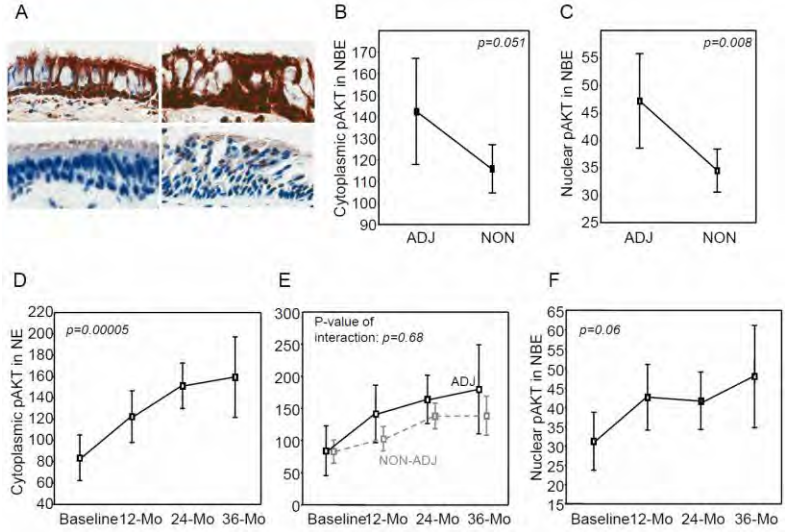

Figure 7. Immunohistochemical analysis of pAKT-Thr308 in airway biopsy specimens of the Vanguard cohort. **A**. Representative photomicrographs of pAKT-thr308 immunostaining (upper, strong; lower, weak). ANOVA analysis of cytoplasmic (**B**) and nuclear (**C**) pAKT-Thr308 immunohistochemical expression in adjacent (ADJ) and non-adjacent (NON) normal bronchial epithelia (NBE). ANOVA analysis of cytoplasmic (**D-E**) and nuclear (**F**) pAKT-Thr308 expression change by time in the Vanguard airway biopsies.

IHC expression of pAKT-Thr308 was compared across the examined airway biopsies based on site from the original location of the tumor as well as time the biopsy was performed following the baseline timepoint. Cytoplasmic pAKT-Thr308 expression exhibited a trend towards an increase in ADJ airways compared to non-adjacent and contralateral airways (NON) (**Figure 7B**). Nuclear pAKT-Thr308 IHC expression was significantly higher in adjacent compared to

non-adjacent and contralateral airways (p=0.008 of ANOVA test) (**Figure 7C**). Interestingly, when we compared and contrasted pAKT-Thr308 expression based on time, ANOVA analysis revealed a significant up-regulation of the cytoplasmic expression of the phosphorylated protein with time in the airway biopsies (p=0.00005, **Figure 7D**). Moreover, the difference in the increase in cytoplasmic pAKT-Thr308 expression with time between adjacent and other airways was not significant following testing for the interaction of site and time variables by ANOVA (**Figure 7E**). Nuclear pAKT-Thr308 expression was also up-regulated with time in the analyzed airway biopsies, but was less significant when compared to cytoplasmic expression of the phosphorylated protein (**Figure 7F**). It is important to note that analysis of pAKT-thr308 was performed in normal bronchial epithelia (NBE) only. Similar results were obtained when biopsies from main carinas were included in the analyses.

B. Gene expression analysis of bronchial epithelial samples obtained from lobectomy specimens from NSCLC patients (*Field Cancerization Study*) (Sub-specific Aim 1A and 1C):

To increase our understanding of the molecular basis of lung cancer pathogenesis, we analyzed the transcriptome profiles of cytological specimens (brushings) obtained at lobectomy or pneumonectomy procedures from 2-5 bronchioles with differential proximity from resected tumors and from resected lung tumors and normal parenchyma (**Figure 8**). Samples were obtained from patients undergoing lobectomy or pneumonectomy procedures (n=23) under an MD Anderson institutional review board (IRB)-approved protocol. A summary of the clinicopathological characteristics of the studied patient cases is depicted in **Table 1**. Tumor and normal parenchyma specimens for transcriptome profiling were collected using three different techniques per patient; brushing, snap-freezing in liquid nitrogen and preservation in RNA*later*. Brushings of tumor and normal parenchyma as well as bronchial brushings from airways were performed using Cytosoft brushes (CardinalHealth, Dublin, OH) and placed in Qiazol (Qiagen, Valencia, CA) and immediately placed in dry ice and stored in -80 °C until RNA isolation. Cytological quality control of

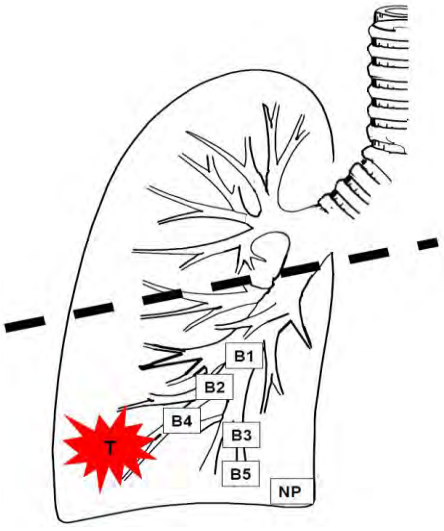

Figure 8. Schematic illustration of the lung airway for molecular mapping of field cancerization by high-throughput profiling. Tumor (T) and normal parenchyma (NP) samples were obtained by brushing, preservation in RNA*later* and snap-freezing in liquid nitrogen. Airway samples (B1 to B5) were obtained by brushing.

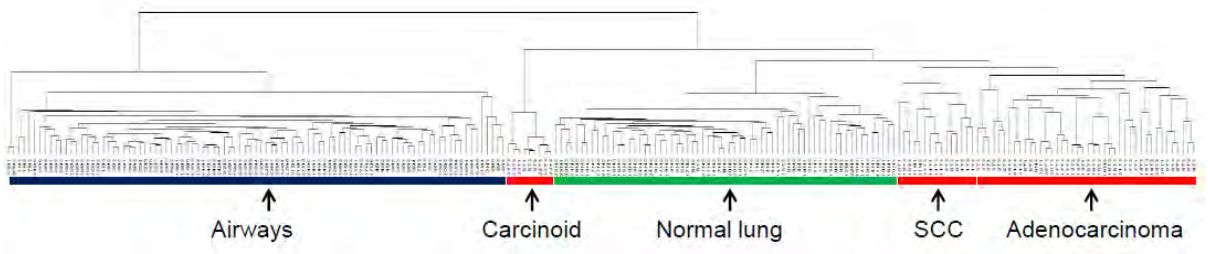

Figure 9. Hierarchical clustering analysis of the lung field cancerization samples. Unsupervised hierarchical clustering analysis of field cancerization samples (n=226) analyzed by microarray profiling. All genes present on the Human Gene 1.0 ST array (Affymetrix) were median centered and arrays clustered by average linkage. Airway (blue), normal lung (green) and tumor (red) major clusters are indicated by the color bars. Tumor sub-clusters (carcinoids, SCCs and adenocarcinomas) are indicated by name underneath the arrows.

the epithelial and malignant cell content of the collected histological tissue specimens was performed and is available. Total RNA of all samples was isolated using the miRNeasy Mini kit following homogenization of tissues and brushing collections. RNA quality was assessed by the 28S/18S ribosomal RNA ratio. 226 samples were eligible for microarray profiling based on RNA quality. Processed RNA samples were hybridized to Affymetrix GeneChip® Human Gene 1.0 ST arrays and were then scanned using the GeneChip® Scanner 3000 from Affymetrix (Santa Clara, CA) to yield raw image files that were subsequently converted to probe set data. Raw expression data was analyzed using the BRB-ArrayTools v.4.1.0 developed by Dr. Richard Simon and the BRB-ArrayTools Development Team and then normalized and log-transformed by the robust multi-array analysis (RMA) method using the R language environment.

We first analyzed the entire set of collected and microarray-profiled field cancerization samples (n=226) by unsupervised hierarchical clustering analysis. The entire set of genes present on the Human Gene 1.0 ST array (n=33,251) were median-centered and samples were clustered by average linkage. The unsupervised clustering revealed the presence of major clusters, as indicated by the colored horizontal bars (airways, blue; normal lung, green; tumors, red) based on histopathology of the specimen analyzed (**Figure 9**). Interestingly, the tumor cluster was further divided into sub-clusters based on the type of lung tumor profiled (carcinoids, SCCs and adenocarcinomas). In addition, all carcinoid tumors (n=3 cases, 9 samples) grouped into an independent sub-cluster and all NSCLCs were divided into two major sub-clusters with one harboring entirely lung SCC cases and the second lung cluster including all lung adenocarcinoma cases as well as a NSCLC (sarcomatoid) case.

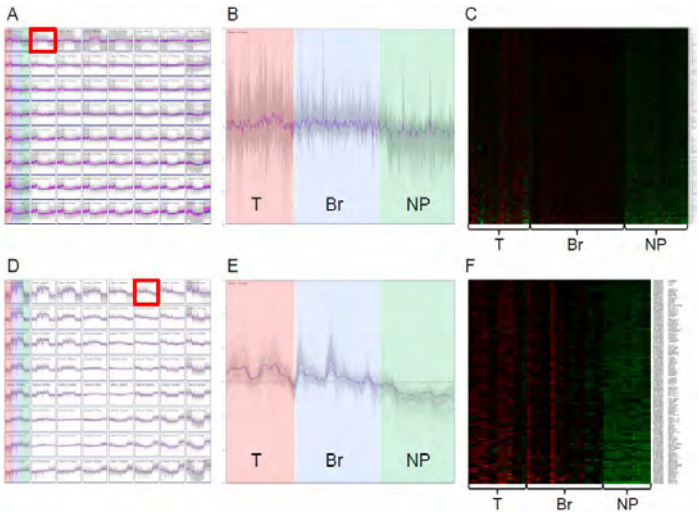

Figure 10. Analysis of differentially expressed genes among lung tumors, airways and normal parenchyma. ANOVA and SOM analyses were performed for all tumors (**A**, **B** and **C**) and SCC cases alone (**D**, **E** and **F**). SOM analysis was to generate 8 x 8 neuronal clusters (**A**, all tumors; **D**, SCCs). Representative clusters are highlighted in **A** and **D** and depicted in **B** and **E**. Median-centered expression levels of genes within the representative clusters are depicted in the heatmaps in **C** and **F**.

It is important to note that, in most cases, tumors and normal parenchyma samples collected by brushing, preservation in RNA*later* and snap-freezing in liquid nitrogen sub-clustered in groups of 3 by patient indicating that differences within profiles of tumors and normal parenchyma collected by the three methods (intra-group differences) are smaller than differences between profiles of tumors and normal parenchyma across patients (inter-group differences).

We then sought to identify genes differentially expressed among tumors, airways (bronchial brushings and normal parenchyma. We applied ANOVA using a p-value of 0.001 as a threshold for statistical significance. Significant genes identified by ANOVA were then queried using self-organizing map (SOM) analysis to identify clusters of genes displaying expression among the groups. SOM analysis was performed using Genesis software developed by Alexander Sturn and Rene Snajder (Graz University of Technology, Graz, Austria) using grids of 8 x 8 neuronal clusters (**Figures 10A and 10D**). SOM analysis gave rise to gene clusters with different variations of expression among tumors, airways and normal lung parenchyma (**Figure 10**). This analysis enabled us to highlight genes with an

increased expression from normal parenchyma to airways and to tumors when all tumor cases (n=23) were analyzed (**Figures 10B and 10C**) or when only lung tumors, airways and normal lung parenchyma corresponding to SCC cases (n=5) were analyzed (**Figures 10E and 10F**). The representative cluster depicted in **Figure 10B** was found to be comprised of 362 genes that displayed increased expression in both lung tumors and airways compared to normal lung when all cases (n=23) were analyzed (**Figures 10B and 10C**). A representative cluster depicted in **Figure 10E** was found to be comprised of 140 genes that displayed increased expression in lung SCCs and corresponding airways compared to matched normal lung (**Figures 10E and 10F**). It is noteworthy that the aforementioned analysis was successful in identifying genes with significant different variations of expression, when applied to sub-groups of the dataset, suggesting that such approaches may be useful in understanding the molecular profiles of the field cancerization phenomenon in relation to histology, e.g. lung adenocarcinoma compared to SCCs. Importantly, our data highlighted genes that are up-regulated in both lung airways and tumors compared to normal lung tissue and, therefore, may play important roles in lung cancer pathogenesis and serve as potential targets for chemoprevention.

Table 1. Clinico-pathological features of NSCLCs cases examined in the Field Cancerization Study

Covariate	Levels	N (%)	Covariate	Levels	N (%)
Gender	Female	11 (47.8%)	T	1	8 (34.8%)
	Male	12 (52.2%)		2	12 (52.2%)
Race	Asian	2 (8.7%)		3	0 (0.0%)
	African-American	4 (17.4%)		4	3 (13.0%)
	Caucasian	17 (73.9%)	N	0	20 (87.0%)
Tobacco history	No	8 (34.8%)		1	2 (8.7%)
	Yes	15 (65.2%)		2	1 (4.3%)
Smoking	Current	6 (26.1%)		3	0 (0.0%)
	Former	9 (39.1%)	M	0	23 (100.0%)
	Never	8 (34.8%)		1	0 (0.0%)
Histology	ADC	14 (63.6%)	Stage	I	17 (73.9%)
	SCC	5 (36.4%)		II	2 (8.7%)
	Carcinoid	3 (13.0%)		III	4 (17.4%)
	NSCLC NOS	1 (4.3%)		IV	0 (0.0%)

ADC, adenocarcinoma; NSCLC, non-small cell lung carcinoma; NOS, non-otherwise specified

The bioinformatic analysis of the aforementioned field cancerization pilot dataset (23 patient cases, 226 samples) is ongoing and is expected to be completed by January 2012 (Table 1). Field cancerization profiles will also be analyzed based on smoking status (e.g. lung adenocarcinoma smokers versus non-smokers). Field cancerization profiles will also be analyzed to potentially unravel genes displaying a site-dependent effect in relation to the original resected primary tumor. In addition, genes differentially expressed among lung tumors, airways and normal lung tissue will be functionally analyzed and topologically organized by pathways and gene-interaction network analyses. Moreover, field cancerization profiles of lung

adenocarcinoma and SCC cases will be compared and contrasted in an attempt to identify genes and cell signaling pathways that may play important roles uniquely between the two major subtypes of NSCLC. Collection of lung tumor, normal parenchyma and airway samples is ongoing to profile a field cancerization set comprised of a larger number of patients.

C. Collection of epithelial samples from both bronchoscopy and lobectomy specimens from patients with lung cancer (Sub-specific Aims 1A and 1C):
M.D. Anderson has collected a complete set of bronchoscopy (nasal, buccal, and 3 bronchial brushes) and lobectomy (tumor, normal parenchyma, and 3-5 bronchial brushes) epithelial samples in 7 patients. Samples from this collection will be used to collaborate with Dr. Avrum Spira (Boston University, Boston, MA) and Dr. Pierre Massion (Vanderbilt University, Vanderbilt, TN), in the beginning of the second year of funding, to analyze miRNA expression profiles of a subset of the aforementioned tumor, airway and normal parenchyma samples by next-generation sequencing (RNA-seq). Global small RNA sequences of tumors, normal parenchyma, localized field cancerization (airways collected from resected specimens) as well as nasal epithelia and main carina and main stem bronchi epithelia (collected by endoscopic bronchoscopy) corresponding to a subset of the SCCs and lung adenocarcinoma cases in **Table 1** will be analyzed by RNA-seq. miRNA profiles and mRNA profiles (**Figures 9 and 10**) analyzed by RNA-seq and microarray technology, respectively, will be statistically tested for significant correlations. Importantly, RNA-seq analysis will enable us to identify potentially novel miRNAs in the airway field of cancerization. Briefly, sequences that align to miRBase are used to quantify the levels of known miRNA per sample. The MirDeep algorithm is then used to identify and quantify the expression level of loci representing potentially novel miRNA species based on transcript structure, the relative abundance of sequences predicted to be contained within the precursor and mature miRNA, the predicted RNA folding of the transcribed locus, and its evolutionary conservation.

Patients characteristics Vanderbilt	
n=25	
Age AVG (STD)	64.8 (10.6)
Gender (% Male)	68.0
PKY AVG (STD)	43.4 (39.8)
% with cancer	71
Histology (%)	
Adeno	61.7
Squamous	31.7
Large cell	5.0
Small cell	8.3
Stage (%)	
I	57.8
II	13.3
III	26.7
IV	2.2

Table 2 Characteristics of patient samples from Vanderbilt.

The Vanderbilt site collected another set of bronchoscopic specimen including nasal, bronchial brushings airway cells associated with lobectomy specimens consisting of normal bronchus and tumor specimens in 25 patients. The samples are currently stored at Vanderbilt and will be used for validation of these signatures. The standardized protocol to allow collection of epithelial brushings of distal airways will be implemented in year 2. The patient's characteristics are described below in Table 2. These samples will be the object of biomarker candidate discovery in Aim 3 of the grant.

Specific Aim 2: To increase our understanding of the role of tumor-initiating stem/progenitor cells in the pathogenesis of lung cancer in the "field cancerization" that develops in current and former smokers.

Summary of Research Findings:

A. The identification of stem cell markers in the airway that are present in premalignant lesions and lung cancer (Sub-specific Aim 2A):
The goal of Specific Aim 2A is to identify markers and profiles of stem cells in the airways and in lung cancer. To this end, we performed validation of antibodies directed against several

biomarkers of cancer stem cells, including Snail, CD44, CD24, and ALDH1A1. We acquired 23 formalin fixed paraffin embedded (FFPE) clinical specimens containing lung squamous cell carcinoma (SCC), adenocarcinoma (ADC), premalignant lesions, and adjacent normal large and small airways. We quantified Snail staining of premalignant lesions versus the relevant normal adjacent tissues and found Snail to be significantly more highly expressed in the premalignant lesions than the normal airway epithelium (**Figure 11**). The CD44 immunostaining results indicate that only the basal layer of the normal bronchial epithelium is positive, while all cells within the squamous metaplasia lesions are CD44 positive (**Figure 12i**). Both the normal bronchial epithelium and the regions of squamous metaplasia appear negative for CD24 staining; while the positive control tonsil tissue stained intensely (**Figure 12ii**). While we detected only cytoplasmic ALDH1A1 staining of a liver positive control tissue, and in normal airway epithelium, a SCC-adjacent region of reserve cell hyperplasia was positive for both cytoplasmic and nuclear ALDH1A1 staining (**Figure 12iii**). Strong expression of Snail, CD44 and ALDH1A1 therefore appear to be markers of premalignant cells in the airway and could be used to identify potential tumor-initiating cells of the airway.

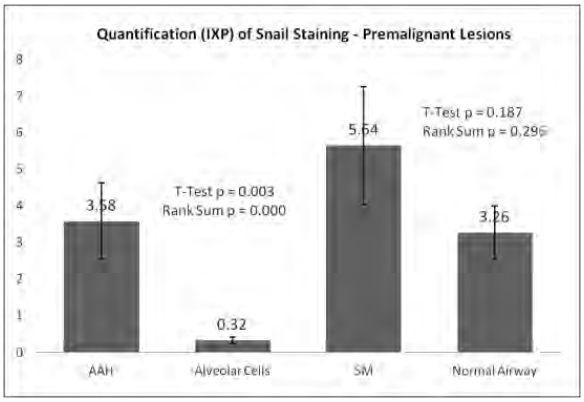

Figure 11. Quantification of Snail expression in human premalignant lung lesions. The number of cases available for evaluation varied for each tissue type under consideration: AAH = 12 cases, SM = 11 cases, alveolar cell areas = 28 cases, and large airways = 27 cases. Tissues were identified and Snail staining was evaluated by our collaborating surgical pathologist (MCF). Intensity (I) grading: 0-3, where 0 was negative and 3 was high. Percent-positive (P) grading: 1-4, where 1=0-25%, 2=26-50%, 3=51-75%, and 4=76-100%. Quantification was based on IXP values. Importantly, truly normal lung tissues from six non-cancer trauma cases were uniformly negative for Snail expression.

HT Genomics qNPA Platform 10 Sample Pilot / 10 Housekeeping Genes / $500			
Specimen Type	**Cell Number**	**Stain Type**	**Region**
FFPE	500	None	AAH Lesion
FFPE	500	H&E	AAH Lesion
FFPE	500	Snail+	AAH Lesion
Frozen	500	None	AAH Lesion
Frozen	500	H&E	AAH Lesion
Frozen	500	Snail+	AAH Lesion
FFPE	1000	None	AAH Lesion
FFPE	1000	H&E	AAH Lesion
FFPE	1000	Snail+	AAH Lesion
Frozen	1000	None	AAH Lesion
Frozen	1000	H&E	AAH Lesion
Frozen	1000	Snail+	AAH Lesion

Table 3. qNPA platform pilot study for cancer stem cell biomarkers in premalignant lesions

Because the Snail antibody has been extensively validated and our results have proven reproducible, we are now utilizing this antibody in a small pilot study to evaluate the ideal conditions and minimum target cell number required for detection of cancer stem cell biomarkers in FFPE clinical specimens via the quantitative nuclease protection assay (qNPA) platform (**Table 3**). The laser capture microdissected (LCM) target cells will be shipped to HTGenomics for evaluation of 10 housekeeping genes. As shown below, we have demonstrated our ability to isolate the atypical adenomatous hyperplasia (AAH) premalignant lesions separately from the normal surrounding type II pneumocytes, which requires single cell selection by LCM (**Figure 13**).

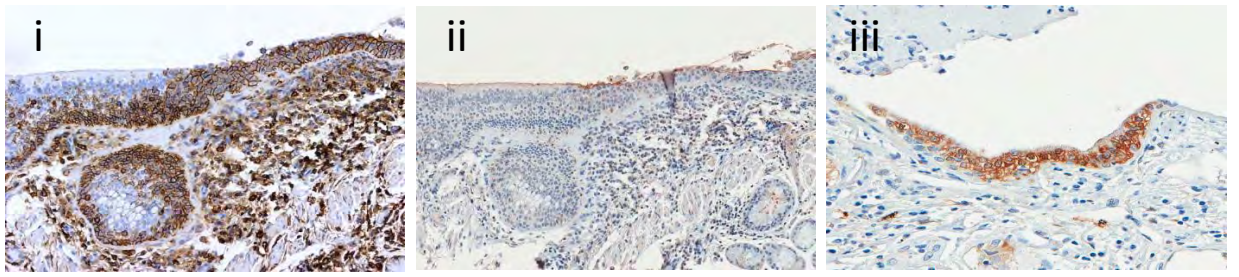

Figure 12. Validation of CD44, CD24, and ALDH biomarkers of cancer stem cells. (Left panel) CD44 staining of a region of squamous metaplasia within a FFPE SCC clinical specimen. (Middle panel) CD24 staining of the same squamous metaplasia lesion. (Right panel) ALDH1A1 staining of a region of reserve cell hyperplasia within a FFPE SCC clinical specimen.

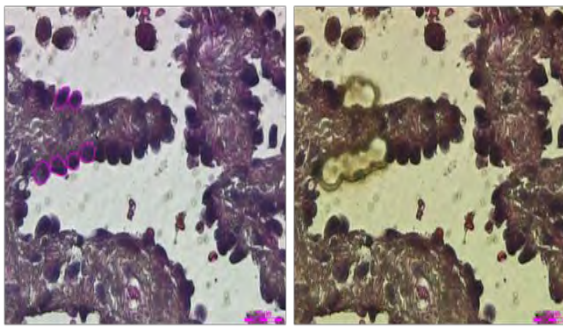

Figure 13. Demonstration of laser capture microdissection isolation of target cells from a premalignant AAH lesion. The Leica LMD7000 was utilized to precisely capture cells of interest from an AAH lesion at the single cell level. Left = Target cell selection; Right = After target cell isolation.

B. Feasibility of sequencing small amounts of RNA from laser captured samples that reflect different pathologic stages of lung carcinogenesis (Sub-specific Aim 2B):

We have made significant technical progress in accomplishing Specific Aim 2B. The UCLA and Vanderbilt groups identified four patients with samples taken at the time of surgical removal of their squamous lung cancer. These samples included normal airway epithelium, premalignant lesions and tumor (**Figure 14**).

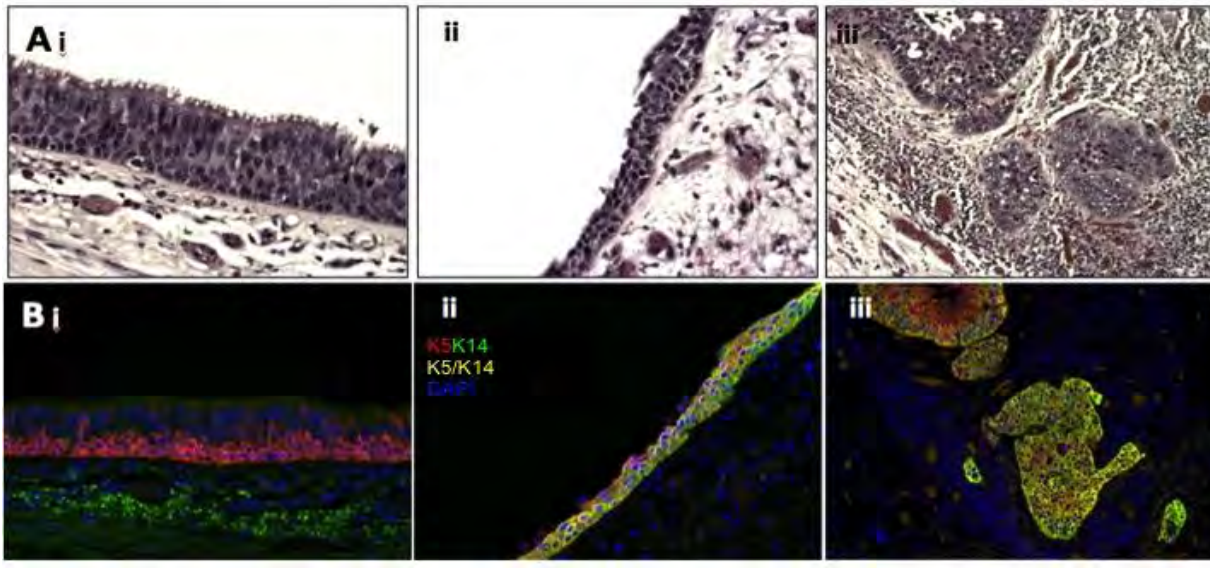

Figure 14. Ai-iii. H&E staining of lesions used for Aim 2B. (i) Normal airway epithelium, (ii) Squamous metaplastic lesion, (iii) Lung squamous cell carcinoma. Bi-iii. K5 and K14 immunofluorescent staining of the corresponding lesions.

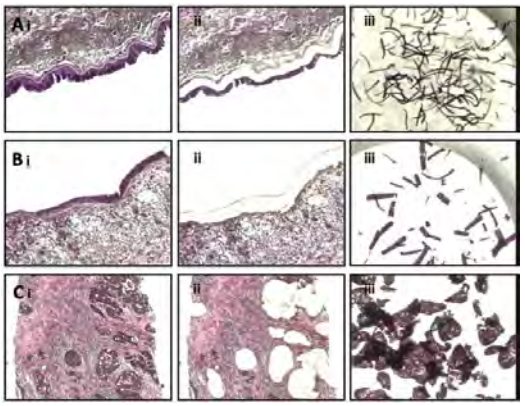

Figure 15. A(i) Normal airway epithelium before laser microdissection, (ii) Normal airway epithelium after laser microdissection of the basal cells, (iii) Collected basal layer of normal airway epithelium. B(i) Squamous metaplastic lesion before laser microdissection, (ii) The same region after laser microdissection of the squamous metaplastic cells, (iii) Collected

We performed laser capture microdissection (LCM) on these tissue blocks to retrieve basal stem cells of the histologically normal adjacent airway epithelium, premalignant lesions and squamous lung cancer all from the same patient (**Figure 15**).

RNA was isolated with the QIAgen RNeasy Micro kit and was found to be present in low amounts and was of low quality as determined by Bioanalyzer analysis (**Figure 16**). The RNA was then successfully converted into cDNA with the NuGEN Ovation RNA-Seq kit and the quality and quantity of cDNA were found to be of sufficient amount and quality to proceed with generating RNA-seq libraries (**Figure 17**). Quantitative real-time PCR was also performed to confirm the expression of KRT5, SOX2, and GAPDH in these libraries (**Figure 18**). The quality of the final libraries generated with the NuGEN Encore Library System was acceptable as determined by Bioanalyzer analysis (**Figure 19**) and the samples were sequenced on Illumina Genome Analyzer IIx or HiSeq machines with read lengths of 36 and 50 base pairs, respectively.

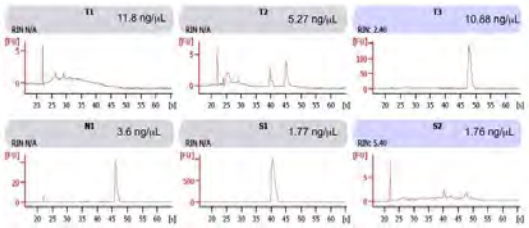

Figure 16. RNA was isolated with the QIAgen RNeasy micro kit, and analyzed on the Bioanalyzer RNA chip. The concentration was measured by nanodrop.

Despite the poor quality of the starting RNA, the sequencing reads were generally of high quality (**Figure 20**), and an average of 46-61% of reads from each patient could be aligned uniquely to the human genome (build human genome hg19) (**Table 3**). The BEDTools utility coverageBed was used to compute reads per kilobase per million (RPKM) values to determine the expression corresponding to 52,974 Ensembl Gene (ENSG) IDs. We confirmed the differential expression of several genes whose expression has previously been reported to be significantly increased or decreased in SCCs (for two examples, see **Figure 21**).

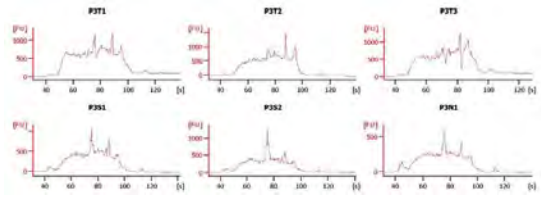

Figure 17. Quality of cDNA generated with NuGEN Ovation RNA-Seq kit was analyzed by the Bioanalyzer high sensitivity DNA chip. The cDNA were in the range of 50-100 bp.

A linear mixed-effects model was used to identify genes whose expression was significantly increased or decreased from normal to premalignant to tumor samples across all four patients (treating sample category as a fixed effect and patient as a random effect). A total of 1210 genes were identified whose expression was associated with tumor progression (940 increased, 270 decreased, $p < 0.01$). An analysis was then performed using

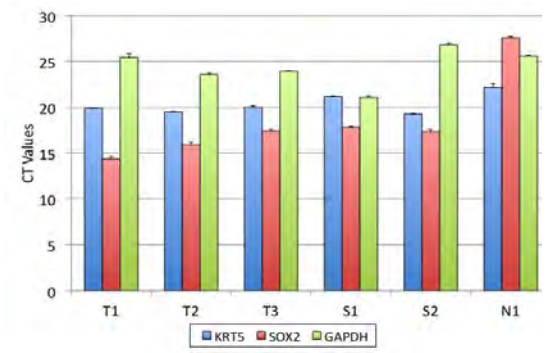

Figure 18. Real-time PCR measuring the transcript levels of KRT5, SOX2, and GAPDH, using the cDNA generated from the NuGEN Ovation RNA-Seq kit as the

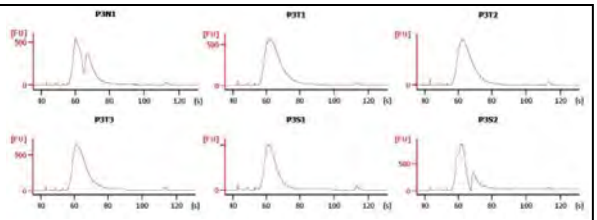

Figure 19. The quality of final libraries generated with NuGEN Encore Library System was tested on the Bioanalyzer high sensitivity DNA chip.

Patient	Platform	Read length	Sample	Total reads	Aligned uniquely	% Aligned
1	GAIIx	36 bases	N1	17,125,822	10,476,503	61%
			N2	15,977,880	8,435,845	53%
			N3	18,281,944	10,622,379	58%
			S1	19,984,712	13,053,687	65%
			T1	19,249,684	11,596,806	60%
			T2	26,594,428	18,477,445	69%
2	HiSeq	50 bases	N1	130,967,326	66,571,270	51%
			S1	137,829,335	59,683,093	43%
			S2	127,287,840	61,316,526	48%
			T1	114,280,863	52,114,221	46%
			T2	165,423,718	73,561,086	44%
			T3	131,976,161	61,500,242	47%
3	HiSeq	50 bases	N1	103,568,098	57,945,824	56%
			N2	131,179,292	64,227,425	49%
			S1	128,758,704	68,887,432	54%
			S2	107,210,554	57,338,119	53%
			T1	77,839,601	33,971,127	44%
			T2	100,645,559	44,343,369	44%
			T3	129,113,308	59,734,659	46%
4	GAIIx	36 bases	N1	33,310,024	11,004,362	33%
			S1	25,274,968	12,625,203	50%
			S2	29,689,509	17,403,581	59%
			T1	30,686,508	15,813,947	52%
			T3	20,785,936	11,088,113	53%

Table 4. Unique alignment of reads to the human genome.

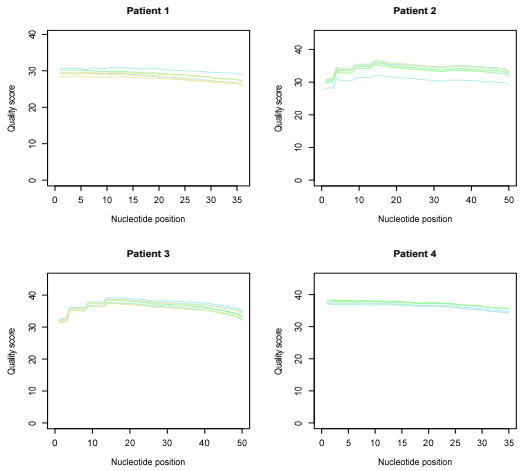

Figure 20. Quality scores of RNA sequencing reads. Within each patient, the mean Phred quality score (-10 * $\log_{10}$(probability of error)) is plotted for each sample at each nucleotide position of the read. Higher scores indicate higher quality, e.g., scores of 30 and 40 indicate error rates of 0.001 and 0.0001, respectively. Normal (N), premalignant (P), and tumor (T) samples are labeled gold, green, and blue, respectively.

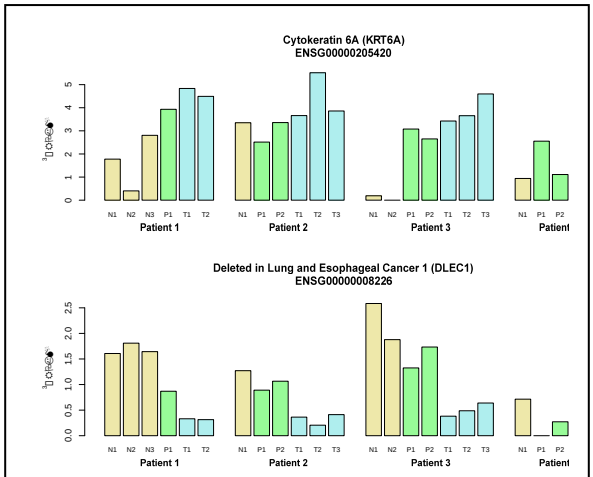

Figure 21. Expression patterns of genes previously reported to be differentially expressed in SCC. RPKM values were obtained for the Ensembl Gene regions corresponding to the cytokeratin KRT6A or the tumor suppressor DLEC1. The cubic root of the RPKM values are plotted for each patient, with normal (N), premalignant (P), and tumor (T) samples colored gold, green, and blue, respectively.

DAVID (http://david.abcc.ncifcrf.gov) to identify GO terms, KEGG pathways, and other terms that were enriched within the top 1000 genes whose expression increased with respect to tumor progression. DAVID analysis identified that terms relevant to cell cycle progression and cell growth were enriched within this list (**Table 5**). A heat map of the expression genes within the KEGG cell cycle gene set hsa04110 is shown in **Figure 22**. The DAVID analysis also identified that the set of the top 1000 genes positively associated with tumor progression is significantly enriched in genes located at cytobands 3q28 and 3q29 (p = 0.02 and 0.0056, respectively). This is in accordance with previous reports of a recurrent amplification of the q26-

Term	Genes	p value	Fold Enrichment
GO 0000278: mitotic cell cycle	49	6.25E-11	2.880298948
GO 0007049: cell cycle	72	1.56E-08	2.017966652
GO 0000280: nuclear division	29	9.64E-07	2.866939491
GO 0009057: macromolecule catabolic process	66	2.01E-06	1.837960237
GO 0007059: chromosome segregation	16	3.48E-06	4.296137509
GO 0051301: cell division	32	1.47E-05	2.359234836
GO 0051439: regulation of ubiquitin-protein ligase activity during mitotic cell cycle	13	8.39E-05	3.982247181
GO 0006412: translation	32	1.34E-04	2.102641319
GO 0051276: chromosome organization	39	9.26E-04	1.748904432
GO 0006323: DNA packaging	15	9.30E-04	2.78835848
GO 0000070: mitotic sister chromatid segregation	8	0.001082694	4.833154698
GO 0007051: spindle organization	8	0.004120307	3.866523758
GO 0006007: glucose catabolic process	9	0.004754211	3.374875263
KEGG hsa04110: Cell cycle	17	0.002144615	2.336351351

Table 5. DAVID analysis identified that terms relevant to cell cycle progression and cell growth were enriched in the dataset.

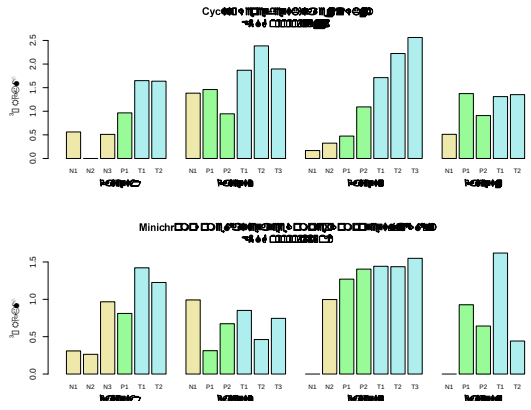

Figure 22. Expression patterns of mitotic cell cycle genes. RPKM values were obtained for the Ensembl Gene regions corresponding to the S-phase regulator MCM7 or the G1-phase regulator CDK4. The cubic root of the RPKM values are plotted for each patient, with normal (N), premalignant (P), and tumor (T) samples colored gold, green, and blue, respectively.

q29 region of chromosome 3 in lung SCCs [1]. A heat map of the expression of the Ensembl genes located between 3q26-3q29 is shown in **Figure 23**.

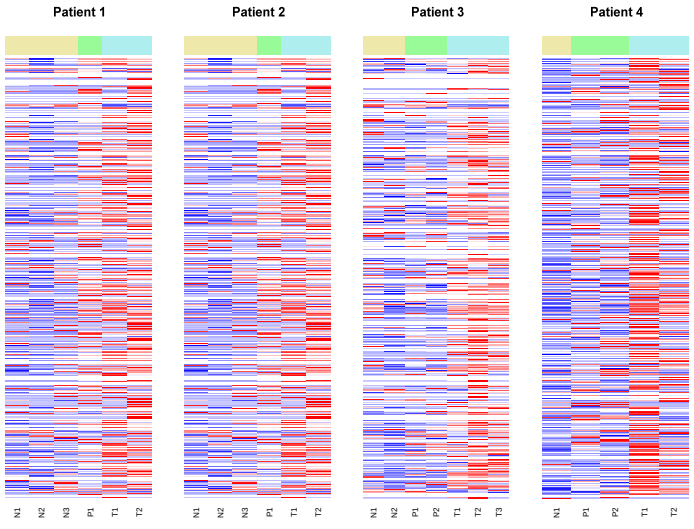

Figure 23. Expression of genes located at chromosome 3q26-3q29. RPKM values were obtained for all Ensembl Gene regions within the chromosome 3q26 to 3q29 (hg19 coordinates 160700000 to 198022430, as reported by the UCSC Table Browser (http://genome.ucsc.edu/cgi-bin/hgTables). Normal (N), premalignant (P), and tumor (T) samples are labeled gold, green, and blue, respectively. Blue and red indicate gene expression that is lower or higher, respectively, than the mean across all samples within a patient.

C. V

alidation of the use of FFPE samples for LCM and RNA-seq (Sub-specific Aim 2B):
Recently we performed LCM on FFPE samples and were able to extract RNA of sufficient quality and quantity from normal basal airway epithelium, premalignant lesions and squamous lung cancer samples to perform RNA sequencing. The quality of the sequencing reads and the proportion of reads aligning uniquely to the genome was similar to those from sequencing of RNA obtained from frozen samples. The ability to use a repository of FFPE samples provides us with a larger pool of samples to work with as we move forward with our studies.

The Vanderbilt group has identified 6 more patient samples with normal bronchial epithelium, premalignant lesions and invasive cancer for this study and is working with the UCLA group to identify more patient samples.

D. Test of feasibility for proteomic studies with *in situ* specimens

The UCLA group identified a molecular profile dominated by Snail that may drive epithelial mesenchymal transition (EMT) and tumor initiating characteristics in the airway epithelium. Snail is over-expressed in human bronchial epithelial cells in premalignant lesions *in situ*. In order to test the feasibility of analyzing these specific cells from *in situ* specimens, we performed preliminary *in vitro* experiments to assess the potential impact of this transcription factor. In that context, we performed Shotgun Proteomic Analysis comparing human bronchial epithelial cells and compared them to the same cells over-expressing Snail.

Cell pellets of each of the epithelial cell tissues were collected at UCLA using the following protocol: a single vial of each cell (1 control and 1 Snail) was grown to confluency, split into separate dishes, grown in parallel and collected at the same time. This procedure has been demonstrated to limit variability of samples due to inconsistent growth and harvesting conditions. Shotgun analysis of these cells was of particular interest to try to elucidate Snail-specific mechanisms to explain previous results obtained by the UCLA group and to assess feasibility to make these discriminations *in situ*.

Each cell pellet was analyzed in duplicate at Vanderbilt using the Jim Ayers Institute standard operating procedure for tissue preparation and analysis. A 0.2 mg protein aliquot was digested and resolved by isoelectric focusing into 15 fractions, which were each analyzed by LC-MS/MS. Thus, there are 6 measurements (2 technical replicates for 3 samples) for the control group and 6 for the Snail+ group. Raw MS/MS data were evaluated using MyriMatch and IDPicker software. Differentially expressed proteins were then identified using Quasi-Tel pair wise comparison.

The dataset overall is good, with 2809 protein groups identified overall (a protein group usually represents a single protein, but sometimes is a small group of indistinguishable proteins with identical peptides). The overall numbers of protein groups in the two sample types are similar (2229 and 2738, respectively for control and Snail+).

Once the identifications were set, expression of all proteins between the sample groups was compared on the basis of spectral counts (the numbers of MS/MS spectra that map to each protein). A model and software called QuasiTel was utilized to fit the count data to a quasi-likelihood model based on a Poisson distribution. The proteins were then sorted using the $\log_2$ Rate Ratio calculated by QuasiTel. Data was also filtered using a multiple comparison-adjusted quasi FDR calculated by the model (this can be thought of as an adjusted P-value).

The following general observations were made:

1. Known markers of EMT (i.e. vimentin) were shown to be overexpressed in the Snail+ cells.
2. Other structural/motility proteins that seem to be consistent with an EMT phenotype were also shown to be overexpressed in the Snail+ cells.

Specific Aim 3: Test airway-based mRNA and microRNA biomarkers of diagnosing lung cancer in current and former smokers at high risk for lung cancer in minimally invasive sites.

Summary of Research Findings:
The studies on this Aim will be carried out in Years 3 and 4 of the grant.

Contributions to the progress report from individual sites:
All the Specific Aims in this proposal are highly collaborative and all individual groups are making progress individually and together. We therefore want to highlight the cohesion and teamwork among our groups. Progress on Specific Aim 1 was largely made by the M.D. Anderson group, together with ongoing collaborative studies between M.D. Anderson, Boston University and Vanderbilt on samples collected from patients at M.D. Anderson, Vanderbilt, Boston University and UCLA.
Progress on Specific Aim 2 was made by a collaboration between the UCLA and Boston University groups as well as the UCLA and Vanderbilt groups. There is also an ongoing collaboration with Vanderbilt and M.D. Anderson.
The close co-operation and collaboration among our groups has allowed us to make significant progress in this first year of funding and will continue to be a priority as we move forward with these studies.

KEY RESEARCH ACCOMPLISHMENTS:

- Identified that gene expression is modulated in a site- and a time-dependent manner in the bronchial epithelium of early stage lung cancer patients.

- Identified several pathways preferentially activated in the airway adjacent to tumors in patients with lung cancer, including those mediated by PI3K, NF-kB and ERK1/2.

- Completed the collection and field cancerization gene expression analysis of 23 patients (n=226 samples) with lung tumors using samples obtained from lobectomy specimens.

- Successful transcriptome analysis of normal airway, premalignant lesions and tumor samples has been performed within four patients.

- Previously reported changes in gene expression between normal airway and SCC (at the level of individual genes, pathways, and chromosomal regions) have been validated in this transcriptome dataset.

- Initial evaluation of the transcriptome data demonstrates that the expression of 1,210 genes was significantly associated with the degree of disease across all four patients ($p < 0.01$).

REPORTABLE OUTCOMES:

- **Manuscript:**

Gomperts BN, Spira A, Massion PP, Walser TC, Wistuba II, Minna, JD and Dubinett SM. Evolving concepts in lung carcinogenesis. *Seminars in Respiratory and Critical Care Medicine*. Semin Respir Crit Care Med. 2011 Feb;32(1):32-43. Epub 2011 Apr 15. PMID: 21500122

- **Abstracts:**

1. Kadara H, Saintigny P, Fan Y, Chow CW, Chu ZM, Lang W, Behrens C, Gold K, Liu D, Lee JJ, Mao L, Kim ES, Hong WK, Wistuba II. Gene expression analysis of field of cancerization in early stage NSCLC patients towards development of biomarkers for personalized prevention. Proceedings of the 102nd Annual Meeting of the American Association for Cancer Research; 2011 Apr 2-6; Orlando, Florida. Philadelphia (PA): AACR; 2011. Abstract #3674.

2. Wistuba I, Kadara H, Kim ES, Hong WK. Molecular Pathology of Lung Cancer & Intermediate Markers of Carcinogenesis. 14th World Conference on Lung Cancer; 2011. Abstract #M19.

CONCLUSIONS:

During our first year of research, we demonstrated a localized field cancerization phenomenon on gene expression in the airway of patients with lung cancer, and we identified several pathways preferentially activated in the airway adjacent to tumors. We will continue to perform sample collections and data analysis of field cancerization specimens obtained from surgically resected lungs from patients with lung cancer to further examine the localized field cancerization phenomenon in the distal airway. In addition, our analysis will allow us to identify genes and cell signaling pathways that may play important roles uniquely between the two major subtypes of NSCLC, adenocarcinomas and squamous cell carcinomas.

In addition, we have identified markers of stem cells in the airway that may represent tumor-initiating cells of the airway and are evaluating profiles of these cells. We have identified Snail as a novel marker of stem cells in the airway that promote EMT. We have made a major technical advance in developing the methods required to use low quality and quantity LCM material for RNA-seq. This allows us to examine the gene expression profiles in premalignancy and compare it to the histologically normal airway epithelium and tumor. We have validated this approach and are analyzing the data to identify novel pathways that might be important in lung carcinogenesis. We have also validated the use of formalin fixed paraffin embedded samples for the LCM RNA-seq studies, which allows us to more easily locate tissues with premalignant lesions. We will therefore use these types of samples as we move forward with the project.

All of these studies are identifying biomarkers that could be used for early lung cancer detection and pathways that are involved in “field cancerization”. Understanding this “field cancerization” and development of premalignant lesions is likely to shed light on novel pathways in lung carcinogenesis that could lead to diagnostic tests, therapies and cancer chemoprevention strategies for lung cancer.

REFERENCES:

1. Hussenet, T., *et al.* SOX2 is an oncogene activated by recurrent 3q26.3 amplifications in human lung squamous cell carcinomas. *PLoS One* **5**, e8960 (2010).

APPENDIX:

Evolving Concepts in Lung Carcinogenesis

Brigitte N. Gomperts, Avrum Spira, Pierre P. Massion, Tonya C. Walser, Ignacio I. Wistuba, John D. Minna, and Steven M. Dubinett

Brigitte N. Gomperts, MD (corresponding author)
Assistant Professor
Department of Medicine
Division of Pulmonary and Critical Care Medicine
David Geffen School of Medicine at UCLA
Department of Pediatrics
Division of Hematology and Oncology
Mattel Children's Hospital at UCLA
UCLA Lung Cancer Research Program
Jonsson Comprehensive Cancer Center
10833 Le Conte Avenue; A2-410 MDCC
Los Angeles, CA 90095, USA
Telephone: (310) 825-6708
E-mail: bgomperts@mednet.ucla.edu

Avrum Spira, MD, MSc
Associate Professor
Department of Medicine
Department of Pathology and Laboratory Medicine
Chief, Section of Computational Biomedicine
Boston University School of Medicine
Director, Translational Bioinformatics Program
Clinical and Translational Scientific Institute
Boston University
The Pulmonary Center
Boston University Medical Center
Boston, MA 02118, USA
Telephone: (617) 414-6980
E-mail: aspira@bu.edu

Pierre P. Massion, MD
Associate Professor
Department of Medicine
Department of Cancer Biology
Thoracic Oncology Center
Vanderbilt-Ingram Cancer Center
Section Chief, Division of Pulmonary and Critical Care Medicine
Veterans Affairs Medical Center
Nashville, TN 37232, USA

Telephone: (615) 322-3412
E-mail: pierre.massion@vanderbilt.edu

Tonya C. Walser, PhD
Adjunct Assistant Professor
Department of Medicine
Division of Pulmonary and Critical Care Medicine
David Geffen School of Medicine at UCLA
UCLA Lung Cancer Research Program
Jonsson Comprehensive Cancer Center
Los Angeles, CA 90095, USA
Telephone: (310) 206-3881
E-mail: twalser@mednet.ucla.edu

Ignacio I. Wistuba, MD
Professor
Department of Pathology
Department of Thoracic/Head and Neck Medical Oncology
Director, Thoracic Molecular Pathology Laboratory
Director, Laboratory Research Program
M. D. Anderson Cancer Center
Houston, Texas 77030, USA
Telephone: (713) 563-9184
E-mail: iiwistuba@mdanderson.org

John D. Minna, MD
Professor
Department of Medicine
Department of Pharmacology
The University of Texas Southwestern Medical Center
Director, Hamon Center for Therapeutic Oncology Research
Director, Molecular Therapeutics of Cancer Program
Harold C. Simmons Cancer Center
Dallas, Texas 75390, USA
Telephone: (214) 648-4900
john.minna@utsouthwestern.edu

Steven M. Dubinett, MD
Professor
Department of Medicine
Department of Pathology and Laboratory Medicine
Department of Molecular and Medical Pharmacology
Chief, Division of Pulmonary and Critical Care Medicine
David Geffen School of Medicine at UCLA
Director, UCLA Lung Cancer Research Program

Jonsson Comprehensive Cancer Center
Los Angeles, CA 90095, USA
Telephone: (310) 267-2725
E-mail: sdubinett@mednet.ucla.edu

Abstract

Lung carcinogenesis is a complex step-wise process that involves the acquisition of genetic mutations and epigenetic changes that alter cellular processes, such as proliferation, differentiation, invasion and metastasis. Here, we review some of the latest concepts in the pathogenesis of lung cancer and highlight the roles of inflammation, the "field of cancerization" and lung cancer stem cells in the initiation of the disease. Furthermore, we review how high throughput genomics, transcriptomics, epigenomics and proteomics are advancing the study of lung carcinogenesis. Finally, we reflect on the potential of current *in vitro* and *in vivo* models of lung carcinogenesis to advance the field and on the areas of investigation where major breakthroughs will lead to the identification of novel chemoprevention strategies and therapies for lung cancer.

Keywords

field of cancerization, inflammation, stem cells, genomics, epigenomics, proteomics

Abbreviations

loss of heterozygosity (LOH), messenger RNA (mRNA), microRNA (miRNA), DNA methyltransferases (DNMTs), non-small cell lung cancer (NSCLC), single-nucleotide polymorphism (SNP), genome-wide association studies (GWAS), liquid chromatography-tandem mass spectrometry (LC-MS/MS), cyclooxygenase 2 (COX-2); prostaglandin E2 (PGE2), epithelial mesenchymal transition (EMT), matrix metalloproteinase (MMP), cancer stem cell

(CSC), bronchoalveolar stem cells (BASC), aldehyde dehydrogenase (ALDH), severe combined immunodeficiency (SCID), epithelial specific antigen (ESA), keratin 5 (K5), keratin 14 (K14), human bronchial epithelial cell (HBEC),

The "field of cancerization"

The "field of cancerization" refers to areas of histologically normal-appearing tissue adjacent to neoplastic lesions that display molecular abnormalities, some of which are the same as those in the tumors [1, 2]. A number of studies, using cytologic and molecular techniques, have established that cigarette smoking creates a field of injury in all airway epithelial cells exposed to the cigarette smoke [2]. Auerbach and colleagues first described the observation of cellular atypia throughout the airways of smokers at autopsy [3], indicating that the cellular injury produced by smoking involves the whole respiratory tract. Recent molecular findings support the stepwise lung carcinogenesis model in which development of this "field of cancerization" with genetically and epigenetically altered cells plays a central role [1, 4-9]. In the initial phase, injury leads to dysregulated repair by stem/progenitor cells, which form a clonal group of indefinitely self-renewing daughter cells. Additional genetic and epigenetic alterations result in proliferation of these cells and expansion of the field, gradually displacing the normal epithelium. Development of an expanding premalignant field appears to be a critical step in lung carcinogenesis that can persist even after smoking cessation.

For example, mutations in *KRAS* have been described in non-malignant histologically normal-appearing lung tissue adjacent to lung tumors [10]. Moreover, loss of heterozygosity (LOH) events are frequent in cells obtained from bronchial brushings of normal and abnormal lungs from

patients undergoing diagnostic bronchoscopy, and they have been detected in cells from both the ipsilateral tumor-containing and contralateral lungs [11]. Likewise, mutations in the *EGFR* oncogene have been reported in normal appearing tissue adjacent to *EGFR* mutant lung adenocarcinoma, *EGFR* mutations occurred at a higher frequency at sites more proximal to the adenocarcinomas than at more distant regions [8, 12]. More recently, global messenger RNA (mRNA) and microRNA (miRNA) expression profiles have been described in the normal-appearing bronchial epithelium of healthy smokers [13, 14], and a cancer-specific gene expression biomarker has been developed in the mainstem bronchus that can distinguish smokers with and without lung cancer [15, 16]. In addition, modulation of global gene expression in the normal bronchial epithelium in healthy smokers is similar in the large and small airways, and the smoking-induced alterations are mirrored in the epithelia of the mainstem bronchus, buccal and nasal cavities [7, 9, 11, 17, 18].

A number of studies from various laboratories have shown that large airway epithelial cells of current and former smokers with and without lung cancer display allelic loss [17, 18], *P53* mutations [5] and changes in promoter methylation [4] and in telomerase activity of non-cancerous epithelial cells [19]. By genome-wide gene expression profiling of a relatively pure population of bronchial airway epithelial cells collected at the time of bronchoscopy, a number of physiologic responses to cigarette smoke exposure have been observed, and many of these changes remain irreversibly altered even after smoking cessation [14, 16]. It has also been shown that gene expression profiles in the cytologically normal bronchial airway epithelium can predict, with high sensitivity and specificity, the presence of lung cancer in current or former smokers being evaluated for clinical suspicion of lung cancer [15]. This 80 probe set combined with clinical risk factors for disease (age,

smoking history, mass size and lymphadenopathy) produces a biomarker with close to 100% negative predictive value and 95% positive predictive value [16].

Profiling the "field of cancerization" with high throughput molecular analyses (Table 1) [20]

i. Epigenomics

Epigenomics refers to high throughput studies of epigenetic changes. Epigenetic alterations are heritable changes in gene expression without alterations in DNA sequence. These changes encompass DNA methylation, histone modifications/chromatin changes and miRNA level alterations, and they play a vital role in the regulation of gene expression.

DNA methylation. DNA methylation at CpG dinucleotides in the 5' region of genes is a major epigenetic mechanism of gene expression regulation [21, 22]. DNA methylation is mediated by DNA methyltransferases (DNMTs). DNMT1, a maintenance DNMT, acts on pre-existing hemi-methylated substrates to maintain methylation patterns after DNA replication [23]. Two other DNMTs, DNMT3a and DNMT3b, act as *de novo* methyltransferases that catalyze the methylation of unmethylated DNA. Importantly, DNMT3a/b may also promote demethylation of DNA at promoters during cyclical demethylation and remethylation related to the transcriptional activity of these genes.

Genomic DNA *hypo*methylation, leading to genomic instability, and aberrant promoter *hyper*methylation, leading to inactivation of tumor suppressor genes [24, 25], have both been shown to be common events in human cancers. Promoter hypermethylation has been detected in the

blood [25], bronchial lavage fluid [23], induced sputum [26] and pleural fluid [27] of lung cancer patients. *TP16* promoter methylation was found in the sputum of smokers up to 3 yrs before their clinical diagnosis of squamous cell carcinoma [28]. Furthermore, methylation of the promoter region of four genes (*TP16*, *CDH13*, *RASSFIA* and *APC*) in patients with stage I non-small cell lung cancer (NSCLC) was associated with early recurrence [29]. High throughput technology is now allowing the identification of novel target genes for aberrant methylation [30, 31]. Protein expression of one of these, *OLIG1*, was found to correlate significantly with survival in lung cancer patients [32].

Histone modifications and chromatin changes. Chromatin structure is critical in the regulation of gene expression, and alterations in its structure have been linked to changes in DNA methylation, histone methylation and acetylation patterns, depending on the target gene. The acetylated state of histones is associated with transcriptional activity, and active histone acetylation has been shown to play a role in re-expression of silenced tumor suppressor genes [33]. Recent studies indicate that histone deacetylase inhibitors have antitumor activity against NSCLC [34-36]. In addition, histone demethylases act to remove methyl groups and have been linked to a number of cellular processes, including DNA repair, replication, transcriptional activation and repression [37].

miRNAs. miRNAs are small non-coding RNA molecules that play important roles in the epigenetic control of diverse cellular processes by altering the translation of proteins from mRNAs. miRNAs have emerged as key post-transcriptional regulators of gene expression, involved in many physiological and pathological processes, such as proliferation, differentiation, death and stress resistance, by altering levels of gene expression [38]. A single miRNA can target

many different mRNAs, and an mRNA can be targeted by multiple miRNAs, thereby creating a complex network of molecular pathways in cells. Interestingly, widespread down-regulation of miRNAs is commonly observed in human cancers and has been linked mechanistically to promotion of cellular transformation and tumorigenesis. More than 50% of miRNA genes are located in cancer-associated genomic regions or in fragile sites, frequently amplified or deleted in human cancer, resulting in frequent copy number alterations, suggesting that differences in miRNA expression may be induced by genomic alterations. Therefore, miRNAs are also suspected to play a role as oncogenes or tumor suppressor genes [39].

In a study analyzing NSCLC and corresponding normal lung tissues, high *hsa-mir-155* and low *hsa-let-7a-2* expression were found to correlate with poor survival in lung adenocarcinomas ($p<0.033$). In another study, low *let-7* expression was also significantly associated with shorter survival ($p<0.0003$), and overexpression of *let-7* in the A549 lung adenocarcinoma cell line inhibited lung cancer cell growth *in vitro* [40]. Subsequently, *KRAS* was shown to be a target of *let-7* [41], and the significance of reduced *let-7* expression in lung carcinogenesis was further supported in studies that showed *let-7* suppressed tumor initiation in an autochthonous NSCLC model of *K-ras*^{G12D} transgenic mice, which was effectively rescued by ectopic expression of *K-ras*^{G12D} lacking the 3′ UTR [42]. *let-7* also inhibited *in vitro* and *in vivo* growth of *K-ras*^{G12D}-expressing murine lung cancer cells and human lung cancer xenografts [43].

In addition to *let-7, miR-17-92* has also been implicated in the pathogenesis and progression of lung cancer, as they both appear to affect the maintenance of "stemness" and cell cycle regulation. In addition to the complex regulatory networks related to miRNAs, other non-coding

RNAs have been found to be important in gene regulation. For example, snoRNA has been demonstrated to have a miRNA-like function [44], and miRNAs may have a novel RNA decoy function [45] . The multiple targets of each miRNA, in addition to the regulatory effects of many non-coding RNAs other than miRNAs, result in extremely complex regulatory networks present in normal and cancer cells. The challenge is to target these regulatory networks to reset the cells to the normal state and remove the regulatory signals associated with the cancerous state.

ii. Genomics and transcriptomics

Genomics refers to high throughput studies of genetic alterations. These technologies use global gene-expression profiles to develop gene signatures that attempt to determine patient prognosis independent of their clinical staging. These technologies have also been used to develop gene signatures that predict the development of lung cancer in high-risk populations and to predict their response to chemotherapy. There are now more than 35 gene signatures that have been published utilizing a mixture of 4 to 133 gene combinations to predict survival, recurrence and metastasis. These signatures were recently reviewed in detail [46]. There is considerable discrepancy in the literature, where many different gene expression profiles with good predictive value for NSCLC are described, but the profiles do not necessarily overlap. This suggests that there may be many biomarkers for predicting outcome and that many of these genes may be functionally important in determining the aggressive behavior of a tumor.

Chromosome abnormalities often correlate with molecular abnormalities and provide a starting point for gene discovery and characterization in the context of a specific disorder. In cancer biology, chromosomal abnormalities carry diagnostic, prognostic and predictive value of

response to treatment. Most solid tumors are genetically unstable and may have losses or gains of whole or large portions of chromosomes, as well as DNA sequence changes of any length attributable to insertion or deletion of the microsatellite one- to four-base DNA repeating units within a tumor [47]. Measures of these genetic variations can also be used to identify novel candidate genes for lung cancer. CGH arrays, based on the high density of bacterial artificial chromosome clones, are used to study genomic copy number variations at high resolution [48-50]. Single-nucleotide polymorphism (SNP) arrays allow accurate measurement of cancer-specific LOH, polymorphisms and copy number variations in a high throughput manner. In lung cancer, amplification of chromosome 3q is one of the most frequent changes observed, and it is also an early event in lung carcinogenesis, as well as in aero-digestive tract tumors [51, 52]. It is found in early stages of lung cancer development, including severe bronchial dysplasia, and is maintained throughout the progression of cancer [53]. In addition, novel high throughput sequencing techniques allow for genome-wide association studies (GWAS) and have been used to identify common low-penetrance alleles influencing NSCLC risk [54] . For example, two SNPs significantly associated with lung cancer risk have been identified in the chromosomal region 15q25.1, the site of *CHRNA3* and *CHRNA5* (nicotinic acetylcholine receptor alpha subunits 3 and 5) and *PMSA4* (proteosome alpha 4 subunit isoform 1), genes that encode protein subunits expressed by airway epithelial cells and known to bind potential lung carcinogens [55]. Two other large genetic epidemiological studies reported very similar results, further suggesting that this genomic region is important in the pathogenesis of lung cancer [56].

Previous work has demonstrated that gene expression profiles of histologically normal bronchial airway epithelial cells collected from smokers and former smokers undergoing medically

indicated bronchoscopy for suspicion of lung cancer are different between patients with lung cancer and those with a benign diagnosis. The expression differences of 80 probe sets can serve as a biomarker that predicts the lung cancer status of independent samples (n=52) with 83% accuracy. This biomarker is considerably more sensitive for detecting early stage lung cancers than bronchoscopy [15]. Importantly, the accuracy of the biomarker is independent of current or cumulative tobacco-smoke exposure and other clinical risk factors for lung cancer [16], suggesting that the biomarker measures some aspect of cancer physiology that is otherwise clinically occult. Consistent with the notion that cancer-specific patterns of gene expression in bronchial airway epithelial cells reflect a carcinogenic process, the PI3K pathway was recently shown to be activated in bronchial airway epithelial cells from patients with lung cancer at both the gene expression level and at the biochemical level [57]. These data suggest that bronchial airway epithelial cells from current and former smokers with lung cancer exhibit cancer-specific properties that can be detected via gene expression profiling and that these can serve as the basis for lung cancer diagnostic biomarkers. Importantly, levels of mRNA do not always correspond with the protein levels due to posttranscriptional modulation of proteins or changes in degradation rates of proteins. It is therefore important to perform proteomic studies in parallel to complement the gene expression data.

iii. Proteomics

Proteomics is the large-scale study of proteins, particularly of their structure and function. Several high throughput technologies have been developed and recently reviewed [20]. Post-translational modifications of proteins, such as phosphorylation, glycosylation and proteolytic processing, are common events that have the potential to significantly modify protein function

and to confer cellular or tissue specificity. Unlike genomic analysis, proteomic analysis has the ability to detect these modifications. In a study using a phosphoproteomic approach based on phosphopeptide immunoprecipitation and analysis by liquid chromatography-tandem mass spectrometry (LC-MS/MS), tyrosine kinases of known oncogenes (e.g., *EGFR* and *c-MET*) implicated in NSCLC carcinogenesis, as well as novel kinases (e.g., *PDGFRa* and *DDR1*), were identified.

Protein signals have been found that allow the classification of lung tumors by histology, the distinction of primary tumors from metastases and the identification of nodal involvement with 75% accuracy. A 15-signal signature has also been developed that can classify patients into good and poor prognostic groups [58]. Specific protein expression patterns have also been associated with areas of normal airway histology, premalignant lesions and invasive lung cancers with about 90% accuracy [59].

1. Inflammation in the pathogenesis of lung cancer

Chronic inflammation in numerous organ sites increases the risk for cancer development to such an extent that inflammation is now considered the "seventh hallmark of cancer" [60]. The link between inflammation and lung carcinogenesis is well established [61, 62]. Cigarette smoke, in particular, is a potent inducer of lung inflammation and plays a key role in lung carcinogenesis [61, 62]. A number of changes are seen in the airways that are associated with chronic inflammation, including alterations in cytokines, chemokines and growth factors released by alveolar macrophages, lymphocytes, neutrophils, endothelial cells and fibroblasts. Inflammation of the airway targets the epithelium for injury, which further drives an abnormal inflammatory

response.

Cyclooxygenase 2 (COX-2). COX-2 is expressed constitutively at low levels in the lung. Its expression is upregulated early after injury in response to cytokines, growth factors and other stimuli, and COX-2 is an important factor in lung carcinogenesis. Cytoplasmic COX-2 expression is upregulated in both adenocarcinomas and squamous lung carcinomas [63], and COX-2 expression has been shown to be higher in lymph node metastases than in the primary tumors [64, 65]. In addition, COX-2 expression in NSCLC has been found to be a poor prognostic indicator [66-68].

Prostaglandin levels are increased by COX-2 during inflammation. Prostaglandins, including prostaglandin E2 (PGE2), are known to promote carcinogenesis [63, 65]. Cytokines, such as IL-1β and TGF-β, and growth factors, including EGF, have been associated with induction of high expression levels of COX-2. Oncogenic events, such as mutant *KRAS* or loss of *P53*, hypoxia and tobacco-specific carcinogens, have also been associated with elevation of COX-2 [63, 69-72]. Persistence of elevated levels of COX-2 in lung cancer cells is associated with loss of IL-10 receptor expression and constitutive nuclear localization of STAT-6 [73, 74]. Elevation of COX-2 and PGE2 levels have been found to promote carcinogenesis by promoting apoptosis resistance [75], proliferation [76], immunosuppression [77], angiogenesis [78], invasion [79] and epithelial mesenchymal transition (EMT) [80].

There is a diversity of prostaglandin receptors that mediate the downstream signaling of prostaglandins. In lung cancer, the effects of COX-2 on PGE2 levels that then act via prostanoid

receptors have been found to be important. The prostanoid receptors are part of the superfamily of G protein-coupled receptors, designated as EP1, EP2, EP3 and EP4. PGE2, and its signaling through the EP4 receptor, has been shown to mediate invasion in NSCLC. Inhibition of COX-2 in tumors has been shown to diminish matrix metalloproteinase (MMP)-2, CD44, and EP4 receptor expression and invasion. These findings indicate that PGE2 regulates COX-2-dependent, CD44- and MMP-2-mediated invasion in NSCLC via EP receptor signaling [64]. Additionally, EP4 receptor blockade and knockdown reduced metastasis in animal models [81]. Thus, blocking the COX-2-dependent PGE2 production or activity by targeting the downstream signaling pathway of COX-2, such as the EP4 receptor, may produce more profound anti-cancer effects than COX-2 inhibition alone.

Epithelial mesenchymal transition (EMT). EMT is the developmental shift from a polarized epithelial phenotype to a highly motile mesenchymal phenotype. While this process is essential and tightly regulated in embryogenesis and development, unregulated EMT is involved in chronic inflammation, fibrosis and cancer progression. EMT results in changes in epithelial proteins, such as E-cadherin, which results in enhanced migration of cells, along with changes in cell shape and adhesion [82]. In addition to the development of metastases, EMT has also been found to regulate early events in carcinogenesis [83]. EMT has been linked to the development of self-renewal properties that are usually associated with stem cells [83].

The link between inflammation and EMT progression in the development of lung cancer and the promotion of resistance to therapy is well recognized [80, 84]. A number of pathways have been found to affect EMT in cancer, e.g., the TGF-β pathway, PI3K/Akt, ROS, receptor tyrosine

kinase/Ras signaling and Wnt pathways [85-87]. Other inflammatory mediators, such as IL-1β and PGE2, up-regulate the zinc-finger E-box–binding transcriptional repressors of E-cadherin, including Snail, Slug and Zeb1, resulting in EMT [80, 88]. COX-2 has also been found to regulate the transcription of E-cadherin in NSCLC, and a reciprocal relationship between COX-2 and E-cadherin, as well as Zeb1 and E-cadherin in NSCLC, has been described [80]. COX-2 and PGE2 overexpression resulted in a significant reduction in E-cadherin expression via a Zeb1 and Snail transcription factor-mediated mechanism, and inhibition of COX-2 resulted in rescue of E-cadherin expression [80].

Immunosuppression. Immunosuppression may contribute to lung carcinogenesis by allowing lung cancer cells to escape immune surveillance. Tumor cells may contribute to immunosuppression by releasing suppressive cytokines, augmenting the trafficking of suppressor cells to the tumor site and/or promoting differentiation of effector lymphocytes to a T-regulatory cell phenotype. One major impediment to effective therapy is our inadequate understanding of how lung cancer cells escape immune surveillance and inhibit anti-tumor immunity. In previous studies, an immune suppressive network in NSCLC that is due to overexpression of tumor COX-2 has been defined. COX-2 metabolites have been identified as mediators of immunosuppression. PGE2 promotes the CD4+CD25+ T regulatory phenotype and increases the expression of the forkhead transcription factor FOXP3 that is known to program the development and function of T-regulatory cells [89, 90].

COX2 and other signaling pathways. Studies have demonstrated that EGFR and COX-2 have related signaling pathways that may interact to regulate cell proliferation, migration and invasion

[91]. PGE2 has been found to completely overcome the growth inhibitory activity of an EGFR tyrosine kinase inhibitor (TKI) in about 40% of NSCLC cell lines [84]. This mechanism of PGE2-induced EGFR-TKI resistance in NSCLC cells lines is mediated through EGFR-independent activation of the MAPK/Erk signaling pathway. Based on these data, there are several ongoing trials assessing COX-2 in combination with TKIs and/or chemotherapy protocols for treatment of lung cancer and for chemoprevention of NSCLC.

2. Lung cancer stem cells

The cancer stem cell (CSC) model of tumor development and progression refers to the presence of a population of rare cells in a tumor that have stem cell properties; namely, they are capable of self-renewal and differentiation into their progeny. In this model, the self-renewal capacity of the CSCs is responsible for maintaining tumor growth indefinitely. Other cells comprising the bulk of the tumor are actively proliferating and differentiating and are, therefore, susceptible to current conventional cancer therapies [92-99]. Consistent with this model, CSCs are considered to be tumor-initiating cells [92-99]. Recently, it was found that CSCs may not necessarily be rare cells within a tumor. Instead, the CSC could be a rare stem cell, a progenitor cell or a differentiated cell that has developed the ability to self-renew [98]. These tumor-initiating cells are thought to arise from cells that have dysregulated repair, resulting in indefinite self-renewal. They are associated with relapse and recurrence of cancers and poor prognosis, presumably due to resistance to chemotherapy and radiotherapy [93-99]. The contribution of CSCs to tumor resistance fits well with the natural history of lung cancer, which is characterized by a high incidence of recurrence and metastasis, leading to the highest mortality rate of all cancers. Classical validation of a CSC tumor-initiating cell population involves reconstituting the human tumor in an

immunodeficient mouse, followed by the indefinite serial xenotransplantation of the CSCs. The following putative CSC populations have been identified in lung cancer **(Table 2)**.

Bronchoalveolar stem cells (BASC). The lung stem cells, termed BASCs, were first described by Kim *et al* [100]. BASCs express markers of both Clara cells (CCSP) and type II pneumocytes (SP-C), are resistant to injury with naphthalene, and proliferate during epithelial repair [100]. BASCs also exhibit self-renewal and are multipotent in clonal assays. Furthermore, BASC expand in response to oncogenic *KRAS* in culture and in precursors of lung tumors *in vivo*. However, the human equivalent of these cells has not yet been isolated, as Sca1+ populations were used in the mouse studies. As a follow up study, Curtis *et al* demonstrated that Sca1+ and Sca1- populations differed in their tumor propagating potential depending on the genotype of the primary tumor from which they were obtained [101].

Aldehyde dehydrogenase (ALDH) and CD133 as biomarkers for lung cancer stem cells. Aldehyde dehydrogenases are a family of intracellular enzymes that are thought to play a role in cellular detoxification, differentiation and drug resistance through the oxidation of cellular aldehydes [102, 103]. Recently, the expression of ALDH proteins have been observed in numerous adult stem cell populations, including hematopoietic and neural stem cells, where they may function to preserve long lived stem and progenitor cells [104-106]. The expression of ALDH enzymes in adult stem cells is also associated with elevated ALDH enzymatic activity and correlates with CD133 expression. Jiang *et al* demonstrated the ability of ALDH expressing cells to serially propagate tumors in nude mice, and determined that they were resistant to chemotherapy [107]. In addition, ALDH expression was associated with poor prognosis in patients

with NSCLC [107]. Eramo *et al* found the CD133 (Prominin-1) surface marker expression in both small cell and non-small cell lung tumors [108]. High numbers of CD133+epCAM+ cells were isolated from fresh lung tumor specimens and were utilized for serial tumor xenografting via subcutaneous injections into severe-combined immunodeficient (SCID) mice. The self-renewal potential of these CD133+ cells remains to be determined, but CD133 expression was found not to be prognostic in NSCLC, although it did correlate with expression of chemotherapy resistance genes [109]. Bertolini *et al* showed that CD133+ cells were associated with increased resistance to chemotherapy and that CD133+/epithelial specific antigen (ESA)+ cells were increased in NSCLC compared with normal lung tissue and had higher tumorigenic potential in SCID mice [110]. Li and colleagues showed that dual expression of CD133 and ABCG2 was an independent predictor of postoperative recurrence for patients with stage I NSCLC and that these tumors had increased angiogenesis [111].

Keratin 14 (K14) as a novel biomarker for lung cancer stem cells. Keratin 5 (K5)-expressing basal cells are considered to be progenitor cells in the adult large airways at steady state and during airway epithelial repair [112-115]. All K14-expressing cells also express K5. Although K14+ progenitor epithelial cells in the airway are important for repair, they are rarely found in the airway epithelium under homeostatic conditions; in contrast, K5+ cells are relatively abundant [113, 114]. K14 expression was found in the repairing airway epithelium, but also in premalignant lesions and a subset of NSCLC tumors [116]. The presence of K14+ progenitor cells in NSCLC tumors after chronic smoking injury was associated with increased mortality from lung cancer [116]. This is consistent with the development of dysregulated repair after injury, leading to a self-renewing K14+ progenitor cell population in premalignant lesions that could survive long

enough to accumulate the genetic and epigenetic changes considered necessary for tumor development [96]. This implicates a novel putative tumor-initiating cell population in a subset of smoking-related NSCLCs with a poor prognosis.

Snail as a novel biomarker for cancer stem cells. Upregulation of Snail and induction of EMT may represent novel signaling events driving lung carcinogenesis. While Snail, Slug, Zeb, and Twist are known to contribute to the progression of established tumors, they are increasingly recognized for their role in neoplastic transformation, as recently reviewed by Sanchez-Garcia [117]. Mani and colleagues were the first to report that induction of EMT in immortalized human mammary epithelial cells leads to acquisition of mesenchymal traits and expression of stem cell markers [118]. More recently, LBX1, which directs expression of Snail and Zeb1, was noted to morphologically transform mammary epithelial cells and to expand the CD44+CD24- cancer stem cell subpopulation [119]. In a study of pancreatic and colon cancers, Zeb promoted tumorigenicity by repressing stemness-inhibiting miRNAs [120]. The role of EMT in acquisition of stem cells characteristics and malignant conversion of the otherwise normal bronchial epithelium is currently being investigated.

In a recent study, squamous cell carcinoma and adenocarcinoma subtypes of NSCLC both overexpressed Snail compared to normal lung tissues[121]. Likewise, premalignant NSCLC lesions over-expressed Snail, often in association with widespread inflammation, as did the proximal and distal airways of chronic obstructive pulmonary disease-involved lungs and premalignant lesions contained therein [122]. These findings suggest the transcription factor is implicated in the earliest pulmonary carcinogenic events.

Expression of stem cell signaling pathway genes as biomarkers for the presence of lung cancer stem cells. The ability of CSCs to self-renew has been attributed to the retention or reactivation of stem cell signaling pathways, such as the Notch, Wnt, and Hedgehog pathways [123]. By capitalizing on the differential expression of self-renewal signaling pathways in lung CSCs, new therapies may be employed to selectively inhibit the self-renewing cancer cell population [124]. For example, the suppression of Notch signaling in breast and brain CSCs resulted in the reduction of self-renewing stem-like tumor cell populations [125-127]. In some lung cancers, the reduction of Notch signaling by gamma-secretase inhibition has been shown to reduce tumorigenicity and colony formation *in vitro*, however, the effect on the lung CSC population has not been determined [128].

3. Conclusions and future perspectives

Many important discoveries related to lung carcinogenesis have been made, but the disease is extremely complex and there are many aspects of the biology that are not well understood. Consequently, the mortality from lung cancer remains higher than that of any other cancer. This review highlights the evolving concept that inflammation in the lungs sets up a field of injury that promotes the development of lung cancer and that the entire epithelium, not just the cancerous region, is involved in the step wise progression to lung cancer. If this is the case, then the injured airway epithelium provides an intriguing site for further investigation and could be targeted via chemoprevention strategies. The revolution in "-omics" approaches will make high throughput studies of this region feasible and hold the key to determining early events in carcinogenesis. Another novel concept is the idea that reparative cells in the field of

cancerization represent tumor-initiating cells, which develop additive and sometimes synergistic molecular changes that result in stepwise progression to lung cancer. The exact populations of tumor-initiating cells and their aberrant signaling pathways remain to be elucidated, as do the specific genetic and epigenetic alterations in these cells that provide the irreversible event for the development of a tumor. Whether these genetic and epigenetic changes in the tumor-initiating cells will be conserved among all individuals or are variable across the population also remains to be determined and will be part of the development of personalized medicine for lung cancer.

Future discoveries in the field of lung carcinogenesis will rely heavily on modeling of the stepwise progression of disease. Currently, the two most important models of the disease are transgenic mouse models and immortalized human bronchial epithelial cell (HBEC) models. In transgenic mice, the somatic activation of *KRAS* has been shown to induce lung adenocarcinomas [129]. Likewise, somatic activation of point mutations of *P53* induced tumors, though *P53* did not, suggesting that point mutant *P53* allels have enhanced oncogenic potential beyond the simple loss of *P53* function [130]. Most importantly perhaps, inactivation of both *KRAS* and *P53* resulted in the development of a mouse model of SCLC [131], which will be extremely valuable for the field.

HBECs are immortalized with *CDK4* and *hTERT* and can be cloned and genetically manipulated, but they do not form colonies in soft agar or tumors in nude mice. HBECs are capable of differentiation into a pseudostratified epithelial layer, similar to that of normal human bronchial epithelium, when grown in an air-liquid interface culture model [132, 133]. This is a useful model system for analyzing the stepwise progression of lung cancer. For example, HBECs manipulated

to have mutant *KRASV12*, *P53* knockdown, or mutant *EGFR*, alone or in various combinations, acquire the ability to grow in soft agar and to invade in three-dimensional organotypic cultures [132].

In summary, we have learned a great deal about the genetic and epigenetic changes that occur after airway injury and are found in lung tumors and the surrounding airway epithelium. Novel technologies will allow us to greatly expand our understanding of the stepwise changes that result in lung cancer and will enable us to identify which cells and molecular changes are responsible for the progression. This is likely to yield important advances for the field where the ultimate goal is development of novel therapies and chemoprevention strategies.

Acknowledgements

This work was supported by DOD CTRA LC090615 to SMD, AS, PPM, IIW, JDM, BNG and TCW and CIRM RN2-009-04 to BNG.

Table 1.

Analyte	**High throughput methods**
Genomics	Whole genome sequencing, CGH arrays, SNP arrays
Epigenomics	miRNA microarrays/sequencing, DNA methylation arrays/sequencing (MeDIP, or bisulfite conversion)
Transcriptomics	RNA-sequencing, gene expression microarrays
Proteomics	2D gel electropheresis, MALDI-TOF MS, Tandem MS, protein arrays, tissue microarrays

Examples of high throughput techniques for molecular analyses.

CGH: comparative genomic hybridization; SNP: single-nucleotide polymorphism; miRNA: microRNA; MeDIP: methylation dependent immunoprecipitation; MALDI-TOF MS: matrix-assisted laser desorption ionization time-offlight mass spectrometry.

Table 2.

Putative Stem Cell (species)	**Location/s in the lung**	**Serial xenografting performed**	**Association with prognosis when present in tumors**	**References**
Broncho-alveolar stem cell (BASC) IF: CC10+SPC+ FACS:Sca1+CD45-Pecam- (mouse)	Bronchoalveolar duct junction	Yes	Not known	Kim et al (2005)
Reparative cell IF: K14+K5+ FACS: N/A (human)	Submucosal gland duct, submucosal glands, repairing airway epithelium, pre-neoplastic lesions	No	P=0.003	Ooi et al (2010)
CD133+ IF and FACS (human)	Not known	Yes	- CD133+ Not significant for prognosis - CD133+ABCG2+ predicts recurrence in Stage I NSCLC p=0.015 -Associated with resistance to chemotherapy	Eramo et al (2008) Bertolini et al (2010) Salnikov et al (2009) Li et al (2010)
ALDH+ (human)	Not known	Yes	P=0.009 -Associated with resistance to chemotherapy	Jiang et al (2009)

Published putative cancer stem cells in lung tumorigenesis.

References

1. Dakubo GD, Jakupciak JP, Birch-Machin MA, Parr RL. Clinical implications and utility of field cancerization. Cancer Cell Int 2007;7:2
2. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 1953;6:963-968
3. Auerbach O, Hammond EC, Kirman D, Garfinkel L. Effects of cigarette smoking on dogs. II. Pulmonary neoplasms. Arch Environ Health 1970;21:754-768
4. Guo M, House MG, Hooker C, et al. Promoter hypermethylation of resected bronchial margins: a field defect of changes? Clin Cancer Res 2004;10:5131-5136
5. Franklin WA, Gazdar AF, Haney J, et al. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. J Clin Invest 1997;100:2133-2137
6. Mao L, Lee JS, Kurie JM, et al. Clonal genetic alterations in the lungs of current and former smokers. J Natl Cancer Inst 1997;89:857-862
7. Steiling K, Ryan J, Brody JS, Spira A. The field of tissue injury in the lung and airway. Cancer Prev Res (Phila Pa) 2008;1:396-403
8. Tang X, Shigematsu H, Bekele BN, et al. EGFR tyrosine kinase domain mutations are detected in histologically normal respiratory epithelium in lung cancer patients. Cancer Res 2005;65:7568-7572
9. Wistuba, II, Lam S, Behrens C, et al. Molecular damage in the bronchial epithelium of current and former smokers. J Natl Cancer Inst 1997;89:1366-1373
10. Nelson MA, Wymer J, Clements N, Jr. Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. Cancer Lett 1996;103:115-121
11. Powell CA, Klares S, O'Connor G, Brody JS. Loss of heterozygosity in epithelial cells obtained by bronchial brushing: clinical utility in lung cancer. Clin Cancer Res 1999;5:2025-2034
12. Tang X, Varella-Garcia M, Xavier AC, et al. Epidermal growth factor receptor abnormalities in the pathogenesis and progression of lung adenocarcinomas. Cancer Prev Res (Phila Pa) 2008;1:192-200
13. Hackett NR, Heguy A, Harvey BG, et al. Variability of antioxidant-related gene expression in the airway epithelium of cigarette smokers. Am J Respir Cell Mol Biol 2003;29:331-343
14. Spira A, Beane J, Shah V, et al. Effects of cigarette smoke on the human airway epithelial cell transcriptome. Proc Natl Acad Sci U S A 2004;101:10143-10148
15. Spira A, Beane JE, Shah V, et al. Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. Nat Med 2007;13:361-366
16. Beane J, Sebastiani P, Whitfield TH, et al. A prediction model for lung cancer diagnosis that integrates genomic and clinical features. Cancer Prev Res (Phila Pa) 2008;1:56-64
17. Bhutani M, Pathak AK, Fan YH, et al. Oral epithelium as a surrogate tissue for assessing smoking-induced molecular alterations in the lungs. Cancer Prev Res (Phila Pa) 2008;1:39-44
18. Sridhar S, Schembri F, Zeskind J, et al. Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. BMC Genomics 2008;9:259
19. Miyazu YM, Miyazawa T, Hiyama K, et al. Telomerase expression in noncancerous bronchial epithelia is a possible marker of early development of lung cancer. Cancer Res 2005;65:9623-9627

20. Ocak S, Sos ML, Thomas RK, Massion PP. High-throughput molecular analysis in lung cancer: insights into biology and potential clinical applications. Eur Respir J 2009;34:489-506
21. Esteller M, Fraga MF, Paz MF, et al. Cancer epigenetics and methylation. Science 2002;297:1807-1808; discussion 1807-1808
22. Ahrendt SA, Chow JT, Xu LH, et al. Molecular detection of tumor cells in bronchoalveolar lavage fluid from patients with early stage lung cancer. J Natl Cancer Inst 1999;91:332-339
23. Cheng X, Blumenthal RM. Mammalian DNA methyltransferases: a structural perspective. Structure 2008;16:341-350
24. Metivier R, Gallais R, Tiffoche C, et al. Cyclical DNA methylation of a transcriptionally active promoter. Nature 2008;452:45-50
25. Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB, Herman JG. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. Cancer Res 1999;59:67-70
26. Belinsky SA, Palmisano WA, Gilliland FD, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. Cancer Res 2002;62:2370-2377
27. Ng CS, Zhang J, Wan S, et al. Tumor p16M is a possible marker of advanced stage in non-small cell lung cancer. J Surg Oncol 2002;79:101-106
28. Palmisano WA, Divine KK, Saccomanno G, et al. Predicting lung cancer by detecting aberrant promoter methylation in sputum. Cancer Res 2000;60:5954-5958
29. Brock MV, Hooker CM, Ota-Machida E, et al. DNA methylation markers and early recurrence in stage I lung cancer. N Engl J Med 2008;358:1118-1128
30. Zhong S, Fields CR, Su N, Pan YX, Robertson KD. Pharmacologic inhibition of epigenetic modifications, coupled with gene expression profiling, reveals novel targets of aberrant DNA methylation and histone deacetylation in lung cancer. Oncogene 2007;26:2621-2634
31. Rauch TA, Zhong X, Wu X, et al. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. Proc Natl Acad Sci U S A 2008;105:252-257
32. Brena RM, Morrison C, Liyanarachchi S, et al. Aberrant DNA methylation of OLIG1, a novel prognostic factor in non-small cell lung cancer. PLoS Med 2007;4:e108
33. Brehm A, Miska EA, McCance DJ, Reid JL, Bannister AJ, Kouzarides T. Retinoblastoma protein recruits histone deacetylase to repress transcription. Nature 1998;391:597-601
34. Mukhopadhyay NK, Weisberg E, Gilchrist D, Bueno R, Sugarbaker DJ, Jaklitsch MT. Effectiveness of trichostatin A as a potential candidate for anticancer therapy in non-small-cell lung cancer. Ann Thorac Surg 2006;81:1034-1042
35. Kim SM, Kee HJ, Eom GH, et al. Characterization of a novel WHSC1-associated SET domain protein with H3K4 and H3K27 methyltransferase activity. Biochem Biophys Res Commun 2006;345:318-323
36. Loprevite M, Tiseo M, Grossi F, et al. In vitro study of CI-994, a histone deacetylase inhibitor, in non-small cell lung cancer cell lines. Oncol Res 2005;15:39-48
37. Kouzarides T. Chromatin modifications and their function. Cell 2007;128:693-705
38. Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. Cell 2003;113:673-676
39. Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell 2006;9:189-198

40. Takamizawa J, Konishi H, Yanagisawa K, et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. Cancer Res 2004;64:3753-3756
41. Johnson SM, Grosshans H, Shingara J, et al. RAS is regulated by the let-7 microRNA family. Cell 2005;120:635-647
42. Kumar MS, Erkeland SJ, Pester RE, et al. Suppression of non-small cell lung tumor development by the let-7 microRNA family. Proc Natl Acad Sci U S A 2008;105:3903-3908
43. Trang P, Medina PP, Wiggins JF, et al. Regression of murine lung tumors by the let-7 microRNA. Oncogene 2010;29:1580-1587
44. Ender C, Krek A, Friedlander MR, et al. A human snoRNA with microRNA-like functions. Mol Cell 2008;32:519-528
45. Eiring AM, Harb JG, Neviani P, et al. miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. Cell 2010;140:652-665
46. Agullo-Ortuno MT, Lopez-Rios F, Paz-Ares L. Lung Cancer Genomic Signatures. J Thorac Oncol 2010
47. Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 1998;58:5248-5257
48. Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 1998;20:207-211
49. Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet 1999;23:41-46
50. Ishkanian AS, Malloff CA, Watson SK, et al. A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet 2004;36:299-303
51. Imoto I, Yuki Y, Sonoda I, et al. Identification of ZASC1 encoding a Kruppel-like zinc finger protein as a novel target for 3q26 amplification in esophageal squamous cell carcinomas. Cancer Res 2003;63:5691-5696
52. Singh B, Gogineni SK, Sacks PG, et al. Molecular cytogenetic characterization of head and neck squamous cell carcinoma and refinement of 3q amplification. Cancer Res 2001;61:4506-4513
53. Levin NA, Brzoska PM, Warnock ML, Gray JW, Christman MF. Identification of novel regions of altered DNA copy number in small cell lung tumors. Genes Chromosomes Cancer 1995;13:175-185
54. Amos CI, Wu X, Broderick P, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. Nat Genet 2008;40:616-622
55. Hung RJ, McKay JD, Gaborieau V, et al. A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. Nature 2008;452:633-637
56. Thorgeirsson TE, Geller F, Sulem P, et al. A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. Nature 2008;452:638-642
57. Gustafson AM, Soldi R, Anderlind C, et al. Airway PI3K pathway activation is an early and reversible event in lung cancer development. Sci Transl Med 2010;2:26ra25
58. Yanagisawa K, Shyr Y, Xu BJ, et al. Proteomic patterns of tumour subsets in non-small-cell lung cancer. Lancet 2003;362:433-439

59. Rahman SM, Shyr Y, Yildiz PB, et al. Proteomic patterns of preinvasive bronchial lesions. Am J Respir Crit Care Med 2005;172:1556-1562
60. Colotta F, Allavena P, Sica A, Garlanda C, Mantovani A. Cancer-related inflammation, the seventh hallmark of cancer: links to genetic instability. Carcinogenesis 2009;30:1073-1081
61. Brody JS, Spira A. State of the art. Chronic obstructive pulmonary disease, inflammation, and lung cancer. Proc Am Thorac Soc 2006;3:535-537
62. Smith CJ, Perfetti TA, King JA. Perspectives on pulmonary inflammation and lung cancer risk in cigarette smokers. Inhal Toxicol 2006;18:667-677
63. Dubinett SM, Sharma S, Huang M, Dohadwala M, Pold M, Mao JT. Cyclooxygenase-2 in lung cancer. Prog Exp Tumor Res 2003;37:138-162
64. Dohadwala M, Batra RK, Luo J, et al. Autocrine/paracrine prostaglandin E2 production by non-small cell lung cancer cells regulates matrix metalloproteinase-2 and CD44 in cyclooxygenase-2-dependent invasion. J Biol Chem 2002;277:50828-50833
65. Lee JM, Mao JT, Krysan K, Dubinett SM. Significance of cyclooxygenase-2 in prognosis, targeted therapy and chemoprevention of NSCLC. Future Oncol 2007;3:149-153
66. Khuri FR, Wu H, Lee JJ, et al. Cyclooxygenase-2 overexpression is a marker of poor prognosis in stage I non-small cell lung cancer. Clin Cancer Res 2001;7:861-867
67. Tsubochi H, Sato N, Hiyama M, et al. Combined analysis of cyclooxygenase-2 expression with p53 and Ki-67 in nonsmall cell lung cancer. Ann Thorac Surg 2006;82:1198-1204
68. Kim HS, Youm HR, Lee JS, Min KW, Chung JH, Park CS. Correlation between cyclooxygenase-2 and tumor angiogenesis in non-small cell lung cancer. Lung Cancer 2003;42:163-170
69. Huang M, Stolina M, Sharma S, et al. Non-small cell lung cancer cyclooxygenase-2-dependent regulation of cytokine balance in lymphocytes and macrophages: up-regulation of interleukin 10 and down-regulation of interleukin 12 production. Cancer Research 1998;58:1208-1216
70. Mao JT, Cui X, Reckamp K, et al. Chemoprevention strategies with cyclooxygenase-2 inhibitors for lung cancer. Clinical Lung Cancer 2005;7:30-39
71. Subbaramaiah K, Altorki N, Chung WJ, Mestre JR, Sampat A, Dannenberg AJ. Inhibition of cyclooxygenase-2 gene expression by p53. J Biol Chem 1999;274:10911-10915
72. Csiki I, Yanagisawa K, Haruki N, et al. Thioredoxin-1 modulates transcription of cyclooxygenase-2 via hypoxia-inducible factor-1alpha in non-small cell lung cancer. Cancer Res 2006;66:143-150
73. Cui X, Zhang L, Luo J, et al. Unphosphorylated STAT6 contributes to constitutive cyclooxygenase-2 expression in human non-small cell lung cancer. Oncogene 2007;26:4253-4260
74. Heuze-Vourc'h N, Zhu L, Krysan K, Batra RK, Sharma S, Dubinett SM. Abnormal interleukin 10Ralpha expression contributes to the maintenance of elevated cyclooxygenase-2 in non-small cell lung cancer cells. Cancer Res. 2003;63:766-770
75. Krysan K, Merchant FH, Zhu L, et al. COX-2-dependent stabilization of survivin in non-small cell lung cancer. FASEB J. 2004;18:206-208
76. Pold M, Krysan K, Pold A, et al. Cyclooxygenase-2 modulates the insulin-like growth factor axis in non-small-cell lung cancer. Cancer Res. 2004;64:6549-6555
77. Sharma S, Stolina M, Yang SC, et al. Tumor cyclooxygenase 2-dependent suppression of dendritic cell function. Clin. Cancer Res. 2003;9:961-968

78. Pold M, Zhu LX, Sharma S, et al. Cyclooxygenase-2-dependent expression of angiogenic CXC chemokines ENA-78/CXC Ligand (CXCL) 5 and interleukin-8/CXCL8 in human non-small cell lung cancer. Cancer Res. 2004;64:1853-1860
79. Dohadwala M, Batra RK, Luo J, et al. Autocrine/paracrine prostaglandin E2 production by non-small cell lung cancer cells regulates matrix metalloproteinase-2 and CD44 in cyclooxygenase-2-dependent invasion. J. Biol. Chem. 2002;277:50828-50833
80. Dohadwala M, Yang S-C, Luo J, et al. Cyclooxygenase-2-dependent regulation of E-cadherin: prostaglandin E(2) induces transcriptional repressors ZEB1 and snail in non-small cell lung cancer. Cancer Res. 2006;66:5338-5345
81. Yang L, Huang Y, Porta R, et al. Host and direct antitumor effects and profound reduction in tumor metastasis with selective EP4 receptor antagonism. Cancer Res 2006;66:9665-9672
82. Lee JM, Dedhar S, Kalluri R, Thompson EW. The epithelial-mesenchymal transition: new insights in signaling, development, and disease. Journal of Cell Biology 2006;172:973-981
83. Mani SA, Guo W, Liao MJ, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 2008;133:704-715
84. Krysan K, Lee JM, Dohadwala M, et al. Inflammation, epithelial to mesenchymal transition, and epidermal growth factor receptor tyrosine kinase inhibitor resistance. J Thorac Oncol 2008;3:107-110
85. Gotzmann J, Mikula M, Eger A, et al. Molecular aspects of epithelial cell plasticity: implications for local tumor invasion and metastasis. Mutation Research 2004;566:9-20
86. Huber MA, Kraut N, Beug H. Molecular requirements for epithelial-mesenchymal transition during tumor progression. Current Opinion in Cell Biology 2005;17:548-558
87. Radisky DC, LaBarge MA. Epithelial-mesenchymal transition and the stem cell phenotype. Cell Stem Cell 2008;2:511-512
88. St John MA, Dohadwala M, Luo J, et al. Proinflammatory mediators upregulate snail in head and neck squamous cell carcinoma. Clinical Cancer Research 2009;15:6018-6027
89. Baratelli F, Lin Y, Zhu L, et al. Prostaglandin E2 induces FOXP3 gene expression and T regulatory cell function in human CD4+ T cells. Journal of Immunology 2005;175:1483-1490
90. Gavin MA, Rasmussen JP, Fontenot JD, et al. Foxp3-dependent programme of regulatory T-cell differentiation. Nature 2007;445:771-775
91. Krysan K, Reckamp KL, Dalwadi H, et al. Prostaglandin E2 activates mitogen-activated protein kinase/Erk pathway signaling and cell proliferation in non-small cell lung cancer cells in an epidermal growth factor receptor-independent manner. Cancer Research 2005;65:6275-6281
92. Ailles LE, Weissman IL. Cancer stem cells in solid tumors. Curr Opin Biotechnol 2007;18:460-466
93. Boman BM, Wicha MS. Cancer stem cells: a step toward the cure. J Clin Oncol 2008;26:2795-2799
94. Clarke MF, Fuller M. Stem cells and cancer: Two faces of eve. Cell 2006;124:1111-1115
95. Huntly BJ, Gilliland DG. Cancer biology: summing up cancer stem cells. Nature 2005;435:1169-1170
96. McDonald SA, Graham TA, Schier S, Wright NA, Alison MR. Stem cells and solid cancers. Virchows Arch 2009;455:1-13

97. Rosen JM, Jordan CT. The increasing complexity of the cancer stem cell paradigm. Science 2009;324:1670-1673
98. Visvader JE, Lindeman GJ. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. Nat Rev Cancer 2008;8:755-768
99. Ward RJ, Dirks PB. Cancer stem cells: at the headwaters of tumor development. Annu Rev Pathol 2007;2:175-189
100. Kim CF, Jackson EL, Woolfenden AE, et al. Identification of bronchioalveolar stem cells in normal lung and lung cancer. Cell 2005;121:823-835
101. Curtis SJ, Sinkevicius KW, Li D, et al. Primary tumor genotype is an important determinant in identification of lung cancer propagating cells. Cell Stem Cell;7:127-133
102. Moreb JS. Aldehyde dehydrogenase as a marker for stem cells. Curr Stem Cell Res Ther 2008;3:237-246
103. Storms RW, Trujillo AP, Springer JB, et al. Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity. Proceedings of the National Academy of Sciences of the United States of America 1999;96:9118-9123
104. Carpentino JE, Hynes MJ, Appelman HD, et al. Aldehyde dehydrogenase-expressing colon stem cells contribute to tumorigenesis in the transition from colitis to cancer. Cancer Res 2009;69:8208-8215
105. Chen YC, Chen YW, Hsu HS, et al. Aldehyde dehydrogenase 1 is a putative marker for cancer stem cells in head and neck squamous cancer. Biochem Biophys Res Commun 2009;385:307-313
106. Huang EH, Hynes MJ, Zhang T, et al. Aldehyde dehydrogenase 1 is a marker for normal and malignant human colonic stem cells (SC) and tracks SC overpopulation during colon tumorigenesis. Cancer Res 2009;69:3382-3389
107. Jiang F, Qiu Q, Khanna A, et al. Aldehyde dehydrogenase 1 is a tumor stem cell-associated marker in lung cancer. Mol Cancer Res 2009;7:330-338
108. Eramo A, Lotti F, Sette G, et al. Identification and expansion of the tumorigenic lung cancer stem cell population. Cell Death and Differentiation 2008;15:504-514
109. Salnikov AV, Gladkich J, Moldenhauer G, Volm M, Mattern J, Herr I. CD133 is indicative for a resistance phenotype but does not represent a prognostic marker for survival of non-small cell lung cancer patients. Int J Cancer 2009
110. Bertolini G, Roz L, Perego P, et al. Highly tumorigenic lung cancer CD133+ cells display stem-like features and are spared by cisplatin treatment. Proc Natl Acad Sci U S A 2009;106:16281-16286
111. Li F, Zeng H, Ying K. The combination of stem cell markers CD133 and ABCG2 predicts relapse in stage I non-small cell lung carcinomas. Med Oncol
112. Engelhardt JF, Schlossberg H, Yankaskas JR, Dudus L. Progenitor cells of the adult human airway involved in submucosal gland development. Development 1995;121:2031-2046
113. Hong KU, Reynolds SD, Watkins S, Fuchs E, Stripp BR. In vivo differentiation potential of tracheal basal cells: evidence for multipotent and unipotent subpopulations. Am J Physiol Lung Cell Mol Physiol 2004;286:L643-649
114. Rock JR, Onaitis MW, Rawlins EL, et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. Proc Natl Acad Sci U S A 2009;106:12771-12775
115. Schoch KG, Lori A, Burns KA, Eldred T, Olsen JC, Randell SH. A subset of mouse tracheal epithelial basal cells generates large colonies in vitro. Am J Physiol Lung Cell Mol Physiol 2004;286:L631-642

116. Ooi AT, Mah V, Nickerson DW, et al. Presence of a putative tumor-initiating progenitor cell population predicts poor prognosis in smokers with non-small cell lung cancer. Cancer Res;70:6639-6648
117. Sanchez-Garcia I. The crossroads of oncogenesis and metastasis. N Engl J Med 2009;360:297-299
118. Mani SA, Guo W, Liao M-J, et al. The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 2008;133:704-715
119. Yu M, Smolen GA, Zhang J, et al. A developmentally regulated inducer of EMT, LBX1, contributes to breast cancer progression. Genes Dev 2009;23:1737-1742
120. Wellner U, Schubert J, Burk UC, et al. The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. Nat Cell Biol 2009
121. Yanagawa J, Walser TC, Zhu LX, et al. Snail promotes CXCR2 ligand-dependent tumor progression in non-small cell lung carcinoma. Clin Cancer Res 2009;15:6820-6829
122. Walser TC, Yanagawa J, Luo J, et al. Snail-induced and EMT-mediated early lung cancer development: Promotion of invasion and expansion of stem cell populations. Proceedings of the Frontiers in Cancer Prevention Research Conference 2008;Abstract PR-11:67
123. Wicha MS, Liu SL, Dontu G. Cancer stem cells: An old idea - A paradigm shift. Cancer Research 2006;66:1883-1890
124. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers--a different disease. Nat Rev Cancer 2007;7:778-790
125. Fan X, Matsui W, Khaki L, et al. Notch pathway inhibition depletes stem-like cells and blocks engraftment in embryonal brain tumors. Cancer Res 2006;66:7445-7452
126. Farnie G, Clarke RB. Mammary stem cells and breast cancer--role of Notch signalling. Stem Cell Rev 2007;3:169-175
127. Konishi J, Kawaguchi KS, Vo H, et al. Gamma-secretase inhibitor prevents Notch3 activation and reduces proliferation in human lung cancers. Cancer Res 2007;67:8051-8057
128. Westhoff B, Colaluca IN, D'Ario G, et al. Alterations of the Notch pathway in lung cancer. Proc Natl Acad Sci U S A 2009;106:22293-22298
129. Meuwissen R, Berns A. Mouse models for human lung cancer. Genes Dev 2005;19:643-664
130. Olive KP, Tuveson DA, Ruhe ZC, et al. Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. Cell 2004;119:847-860
131. Hoffman JA, Giraudo E, Singh M, et al. Progressive vascular changes in a transgenic mouse model of squamous cell carcinoma. Cancer Cell 2003;4:383-391
132. Sato M, Vaughan MB, Girard L, et al. Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. Cancer Res 2006;66:2116-2128
133. Vaughan MB, Ramirez RD, Wright WE, Minna JD, Shay JW. A three-dimensional model of differentiation of immortalized human bronchial epithelial cells. Differentiation 2006;74:141-148

Abstracts:

1. AACR International Meeting 2011

Gene expression analysis of field of cancerization in early stage NSCLC patients towards development of biomarkers for personalized prevention

Humam Kadara[1], Pierre Saintigny[1], Youhong Fan[1], Chi-Wan Chow[1], ZuoMing Chu[1], Wenhua Lang[1], Carmen Behrens[1], Kathryn Gold[1], Diane Liu[1], J. Jack Lee[1], Li Mao[2], Edward S. Kim[1], Waun K. Hong[1], Ignacio I. Wistuba[1].
[1]MD Anderson Cancer Center, Houston, TX; [2]University of Maryland, Baltimore, MD

Background: The identification of early stage non-small cell lung cancer (ES NSCLC) patients (pts) at higher risk for recurrence or second primary tumor (SPT) development is vital to personalizing prevention and therapy. We sought to decipher spatial and temporal patterns of gene expression in the airway field of ever-smoker ES NSCLC pts to better understand lung cancer pathogenesis and predict recurrence or SPT development.
Methods: Pts on the prospective Vanguard study had definitively treated ES (I/II) NSCLC, were current/former smokers, and had bronchoscopies with brushings obtained from the main carina (MC) at baseline, 12, and 24 months following resective surgery and from different anatomical regions at baseline. Expression profiling is ongoing for all eligible pts (41 pts, 326 samples). To query temporal and spatial airway expression profiles, two sets of six pts were selected based on complete processed time point and baseline airway site (3 different sites per pt) arrays (Affymetrix Human Gene 1.0 ST), respectively. Temporally and spatially differentially expressed genes were independently identified based on a $p<0.01$ of a univariate t-test with estimation of the false discovery rate (FDR), studied by hierarchical clustering and principal component analysis (PCA), and functionally analyzed using network analysis.
Results: 871 gene features were differentially expressed among MCs of six NSCLC pts at baseline, 12 and 24 months and were shown to separately group the MCs as evident in both cluster and PC analyses. Moreover, pathways analysis of the temporally modulated genes showed that a gene-network mediated by extracellular regulated kinase (ERK1/2) was most significantly elevated ($p<0.001$) in function between MCs at 24 months versus baseline. 763 and 931 gene features were differentially expressed between MCs and adjacent-to-resected tumors (ADJ) airways and between MC, ADJ and non-adjacent (distant-to-resected tumor) (NONADJ) airways, respectively. Moreover, pathways analysis of the spatially modulated genes revealed that gene-networks mediated by nuclear factor-κB (NF-κB) and ERK1/2-mediated were most significantly elevated ($p<0.001$) in function in ADJ airway samples versus MCs. Furthermore, PCA revealed that while ADJ airway samples grouped separately and closely together, one MC and 3 NON-ADJ airway samples resided closely with ADJ samples, which were then found to originate from 3 pts with evidence of recurrence, SPT or suspicion of recurrence.
Conclusions: Our findings highlight expression signatures and pathways (ERK1/2 and NF-κB) in a “cancerization field” that may drive lung cancer pathogenesis and be associated with recurrence or SPT development in ES NSCLC pts and thus useful for derivation of biomarkers to guide personalized prevention strategies.

Supported by DoD grants W81XWH-04-1-0142 and W81XWH-10-1-1007.

2. World Conference on Lung Cancer 2011

Molecular Pathology of Lung Cancer & Intermediate Markers of Carcinogenesis

I. Wistuba, H. Kadara, E.S. Kim, W.K. Hong

Lung cancer continues to be the leading cause of cancer-related deaths worldwide with over one million deaths each year. Lung cancer mortality is high in part because most cancers are diagnosed after regional or distant spread of the disease had already occurred and due to the lack of reliable biomarkers for early detection and risk assessment. The identification of new effective early biomarkers will undoubtedly improve clinical management of lung cancer and is tightly linked to better understanding of the molecular events associated with the development and progression of the disease. It has been suggested that histologically normal-appearing tissue adjacent to neoplastic lesions display molecular abnormalities some of which are in common with those in the tumors. This phenomenon, coined field of cancerization, has been shown to be important in lung cancer. We have demonstrated than mutations in *EGFR* occur in normal appearing bronchial epithelium adjacent to *EGFR* mutant lung adenocarcinomas in never smoker patients, and also occurred at a higher frequency at sites more proximal to the tumors than at more distant regions. More recently, gene methylation patterns, as well as global mRNA and microRNA (miRNA) expression profiles, have been described in the normal-appearing bronchial epithelium of healthy smokers. Importantly, modulation of global gene expression in the normal bronchial epithelium in healthy smokers is similar in the large and small airways, and the smoking-induced alterations are mirrored in the epithelia of the mainstem bronchus, buccal and nasal cavities. Increasing our understanding of early phases in lung cancer pathogenesis will aid in the identification of early stage non-small cell lung cancer (NSCLC) patients at higher risk for recurrence or second primary tumor development.

Recently, we have performed global gene expression analysis of the field of cancerization in smoker patients with early stages NSCLC to better understand lung cancer pathogenesis and predict recurrence or second primary tumor development. Our findings highlight expression signatures and activation of cancer-related pathways in a cancerization field that may drive lung cancer pathogenesis and be associated with recurrence or second primary tumor development in NSCLC patients. We propose that the study of field cancerization phenomenon using high-throughput molecular profiling (protein, mRNA, miRNA and DNA) methodologies combined with coupled with functional pathway analysis and studies of in vitro and in vivo models will provide promising markers to improve risk assessment, new targets for novel targeted chemopreventive agents, and selecting NSCLC patients who may benefit from chemopreventive interventions to prevent disease recurrence.

Supported by DoD grants W81XWH-04-1-0142 and W81XWH-10-1-1007

TABLE OF CONTENTS

INTRODUCTION

Lung cancer continues to be the leading cause of cancer-related death in both men and women in the United States. The majority of lung cancers are non-small cell lung cancers (NSCLCs) that include squamous cell carcinomas (SCCs) and adenocarcinomas. Lung cancer mortality is high in part because most cancers are diagnosed after regional or distant spread of the disease had already occurred and due to the lack of reliable biomarkers for early detection and risk assessment. The identification of new effective early biomarkers will improve clinical management of lung cancer and is linked to better understanding of the molecular events associated with the development and progression of the disease.

It has been suggested that histologically normal-appearing tissue adjacent to neoplastic lesions display molecular abnormalities some of which are in common with those in the tumors. This phenomenon, termed field of cancerization, was later shown to be evident in various epithelial cell malignancies, including lung cancer. Loss of heterozygosity (LOH) events are frequent in cells obtained from bronchial brushings of normal and abnormal lungs from patients undergoing diagnostic bronchoscopy and were detected in cells from the ipsilateral and contralateral lungs. More recently, global mRNA and microRNA (miRNA) expression profiles have been described in the normal-appearing bronchial epithelium of health smokers. In addition, modulation of global gene expression in the normal epithelium in health smokers is similar in the large and small airways and the smoking-induced alterations are mirrored in the epithelia of the mainstem bronchus, buccal and nasal cavities.

In this program, high-throughput microarray mRNA and miRNA expression analyses will be performed on cytological specimens (brushings) obtained at intraoperative bionchoscopy from the main carina and main ipsilateral bronchus, as well as on specimens obtained at lobectomy procedures from the main lobe bronchus (adjacent to SCCs), sub-segmental bronchus (adjacent to adenocarcinomas) and from the resected NSCLC tumors. We will compare and contrast global gene expression, both mRNA and miRNA, patterns across all the specimens from the entire field and corresponding NSCLC tumors. We seek to derive lung adenocarcinoma and SCC field cancerization signatures signifying the differential mRNA and miRNA expression patterns between the carina and the subsegmental bronchus and main lobe bronchus, respectively. In addition, similar expression profiles between the carina and resected NSCLC tumors will be integrated with available gene expression data of bronchial brushings from the main carina isolated at various time points post-surgery from 40 NSCLC patients; Department of Defense (DoD) VITAL patients. Lastly, functional pathway analysis will be performed to organize differentially expressed genes into topological biological networks in association with miRNA expression. Promising markers derived from this innovative study will be validated at the mRNA and protein level in histological tissue specimens and may be tested in future studies to determine their potential role in improving risk assessment, providing new targets for novel chemopreventive agents and selecting NSCLC patients who may benefit from chemopreventive interventions to prevent disease recurrence.

This report details the progress made during the first year of research. This report will also be included in the comprehensive annual report to be prepared and submitted by Dr. Steven Dubinett as the overall consortium PI.

PROGRESS REPORT

Molecular Profiles for Lung Cancer Pathogenesis and Detection in U.S. Veterans.

Specific Aim 1: **Identify differential mRNA and microRNA expression patterns among various cytological specimens isolated at different levels from the field cancerization and from corresponding NSCLC tumors.**

Summary of Research Findings

A. Gene and protein expression analysis of bronchial epithelial samples obtained from bronchoscopy from NSCLC patients: This analysis was performed on samples obtained from patients enrolled to the Vanguard clinical trial supported by the DoD VITAL grant (W81XWH-04-1-0412, PI Dr. W.K. Hong), and the gene profiling work and analysis was partially supported by the Lung Cancer Consortium grant reported here.

Gene Expression Analysis: Patients on the prospective Vanguard study had definitively treated ES (I/II) NSCLC and were current or former smokers. Patients had bronchoscopies with brushings obtained from the main carina (MC), airways adjacent (ADJ) to the previously resected tumor and from airways distant from the tumor in the ipsilateral (NON-ADJ) and contralateral (CONTRA) lung at baseline, 12, 24 and 36 months following resective surgery (**Figure 1**). Nineteen patients were selected for the study based on airway sampling of at least five different regions per time point and continuously up to 24 or 36 months (391 airway samples from nineteen patients). Total RNA was isolated from brushings using the RNeasy Mini kit (Qiagen) according to the manufacturer's instructions. Due to the paucity of the material, we employed the Nugen WT-Ovation system for RNA amplification (Nugen Technologies, San Carlos, CA). Synthesis of single-stranded DNA, fragmentation and biotin labeling was performed using the WT-Ovation Pico RNA amplification system, WT-Ovation Exon-Module and Encore™ Biotin Module (NuGEN), respectively, according to the manufacturer's instructions. 2-2.5 micrograms of biotin-labeled DNA was then hybridized to Affymetrix Human Gene 1.0 ST arrays from Affymetrix. Analysis began by construction of a mixed-effects model that incorporated information on the site and time (continuous) the bronchial brushing was obtained as fixed effects. Characteristics of fixed effects and their interaction, in terms of number of genes differentially affected by the effect, were examined by fitting beta-uniform-models (BUMs) based on p-value distributions of all genes according to the fixed-effect (**Figure 2**). Histograms of p-value distribution of fixed effects (**Figures 2A, 2B and 2C**) and a smoothened scatter plot of transformed p-values of both site and time fixed effects (**Figure 2D**) demonstrate that there is little evidence for any interaction between site and time. These findings suggest that, although the airway brushings were collected at different sites and at time points following respective surgery from each patient, all

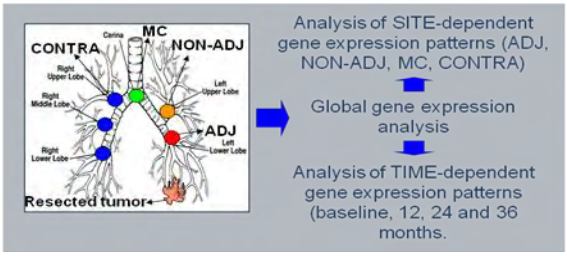

Figure 1. Schematic illustrating sampling and collection of bronchial brushes from different airway sites at different time points following resective surgery from early stage smoker NSCLC patients enrolled in the phase II cancer surveillance Vanguard study.

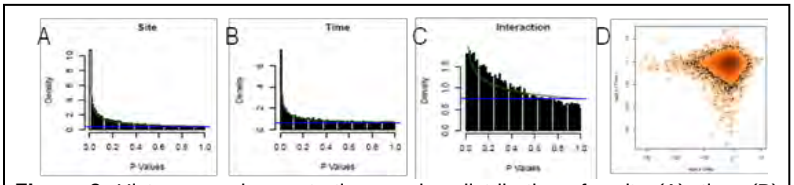

Figure 2. Histograms demonstrating p-value distributions for site (A), time (B) and interaction between both after fitting beta-uniform-mixture (BUM) models (3). D. Smoothed scatter plot of transformed p-values from site and time.

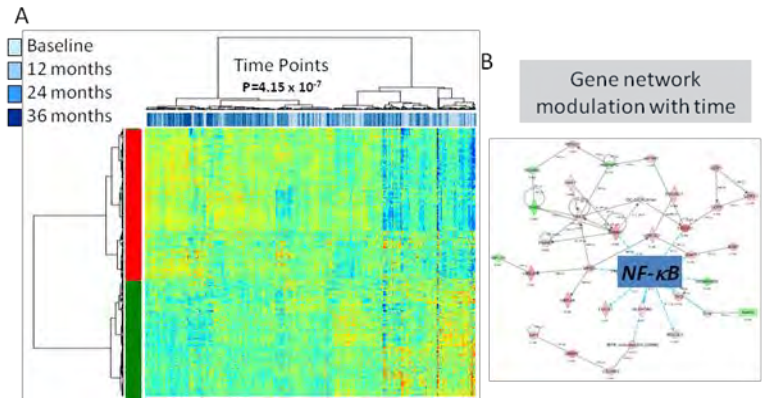

Figure 3. Gene expression profiles are different in a time-dependent manner in the field of cancerization. Hierarchical clustering (A) and functional gene-interaction network analysis (B) of genes (n=1395) differentially expressed continuously with time.

airway samples (n=391) can be utilized independently to unravel genes differentially expressed by site from original resected tumor and by time following removal of the same tumor in all patients. Subsequently, time and site-dependent field of cancerization differential expression profiles were identified based on a false discovery rate (FDR) cut-off of 5% and 1% based on p-value distributions, respectively, studied by hierarchical clustering and functionally analyzed using network analysis. 1,395 genes were determined to be differentially expressed with time in the cancerization fields.

Hierarchical clustering analysis using these genes demonstrated that samples (n=391) were divided into two clusters or branches which were significantly unbalanced with respect to time with the majority of the baseline and 36 months samples separated ($p<0.001$ of the Fisher's Exact test) (**Figure 3A**). Moreover, functional analysis of these genes showed that a nuclear factor-κB (NF-κB)-mediated gene-network was most significantly elevated ($p<0.001$) with time (**Figure 3B**).

1,165 gene features were differentially expressed by site. Two-dimensional clustering of these genes and samples showed distinct classes of differential expression and revealed two main sample clusters with significant separation of ADJ samples from MCs and non-adjacent CONTRA airway samples ($p=0.003$) (**Figure 4A**). Similar results were obtained when the main carina samples were excluded from analysis of genes differentially expressed by site in relation to the original resected tumor (**Figure 4B**). Using both site-dependent analyses, genes were identified that exhibited highest average expression in airways adjacent to tumors (cluster of genes highlighted by red bars) (**Figures 4A and 4B**). It is worthwhile to mention that, following two-dimensional clustering of the site-dependent differentially expressed genes and airway samples, adjacent airways isolated from patients with recurrence or suspicion of recurrence, (black, recurrence; grey, suspicion of recurrence), exhibited on average elevated expression of

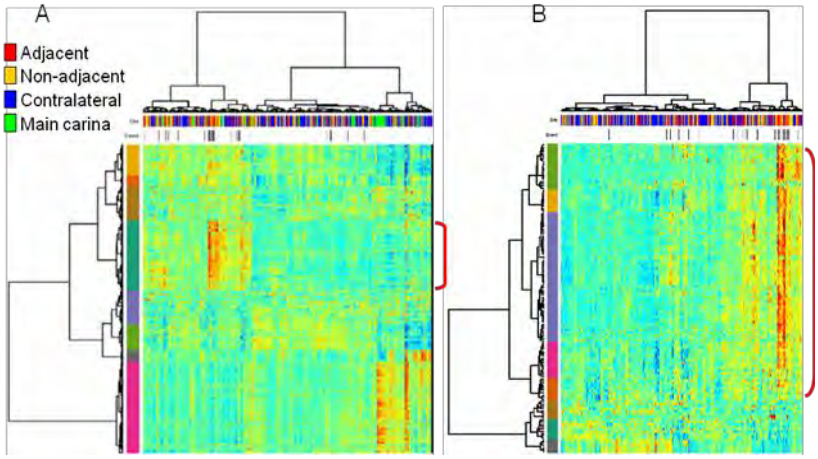

Figure 4. Hierarchical clustering of all samples (**A**) or excluding main carinas (**B**) using genes differentially expressed by site. Recurrences (black) and suspicions (grey) are labeled based on adjacent airways. Genes with highest expression in adjacent airways are indicated by the red bars.

the highest-in-adjacent airway gene cluster compared to adjacent airways isolated from patients with no events in recurrence (**Figures 4A and 4B**). These findings suggest that differential gene expression patterns, by site from original tumor, in the field of cancerization of early stage patients may be associated with recurrence or second primary tumor development.

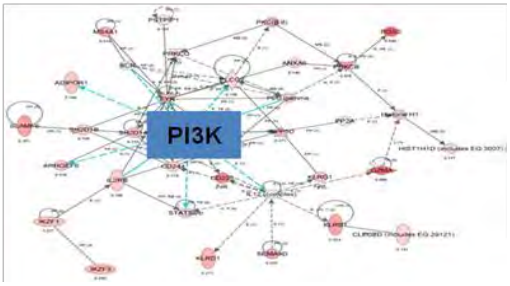

Figure 5. Network analysis of cluster of genes depicted in **Figure 4** with highest average expression in airways adjacent to original resected tumors. PI3K-mediated network is shown as most significantly up-regulated in adjacent airways . red, up-regulated in adjacent airways relative to non-adjacent or contralateral airways.

To further understand the relevance of this gene cluster with highest expression in adjacent airways and isolated from patients with recurrence, we performed functional pathway analysis of the genes using the knowledge-dependent analysis, Ingenuity Pathways analysis (IPA). Functional analysis of the highest-in-adjacent airway genes revealed that gene-networks mediated by PI3K, NF-κB, and ERK1/2 had significantly elevated ($p<0.001$) function in ADJ airway samples, with a gene-interaction network mediated by PI3K being most significantly elevated in function, as predicted by the IPA software, based on number of genes differentially expressed within the network and topological interactions among the same genes (**Figure 5**). These findings suggest that the aforementioned canonical cell signaling pathways, in particular the PI3K survival pathway, may be highly relevant biologically to the molecular pathogenesis of NSCLC, and clinically to predict recurrence or second primary tumor development in early stage patients definitively cured by resective surgery.

We further identified genes differentially expressed by site using different statistical methods. We identified site-dependent gene expression patterns using paired t-test analysis of the 19 NSCLC patients based on differences in expression between adjacent and non-adjacent/contralateral airways. An average ADJ expression score and average non-adjacent/contralateral (NON-ADJ) score was measured for each gene based on all available airway samples per patient. Hierarchical clustering (**Figure 6A**) and principal component (**Figure 6B**) analysis of patients (n=19) based on genes with significant expression differences (by paired t-tests) between ADJ and NON-ADJ samples revealed two main clusters with three of four relapses located in one sub-cluster suggesting potential associations between field of cancerization expression profiles and lung cancer relapse.

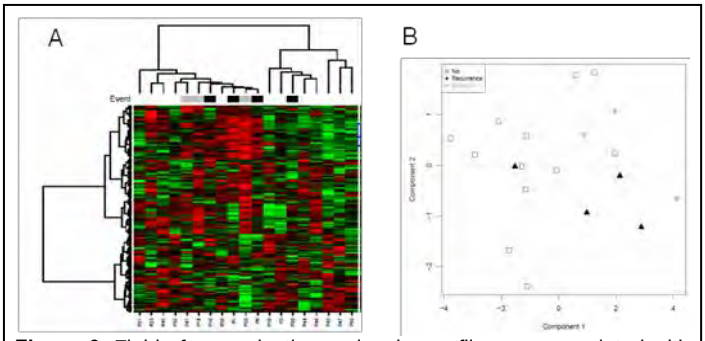

Figure 6. Field of cancerization molecular profiles are associated with patterns of recurrence. Genes with significant expression differences between adjacent and non-adjacent airways (ADJ/NON-ADJ) were used to analyze the 19 pts by clustering (A) and principal component analysis (B). Black: recurrence; grey: suspicion of recurrence.

Protein Expression Analysis: Our findings on the significant modulation of a PI3K-mediated gene-interaction network in adjacent airways compared to other airway brushings (**Figure 5**) prompted us to examine the immunohistochemical (IHC) expression of phosphoAKT (Threonine 308) in available airway biopsy specimens corresponding to the bronchial brushings; we previously analyzed the transcriptome by microarray profiling. AKT is phosphorylated at two major sites, serine 473 and threonine 308, the latter site being phosphorylated through PDK1

following PI3K activation, thus acting as a key surrogate for this pathway activation. We assessed the expression of phosphoAKT-Thr308 by IHC analysis in available and eligible airway biopsy specimens (n=324). Antigen retrieval was performed using the Dako target retrieval system at a PH of 6. Intrinsic peroxidase activity was blocked by 3%methanol and hydrogen peroxide for 15 min and serum-free protein block (Dako) was used for 7 min for blocking non-specific antibody binding. Slides were then incubated at room temperature for 90 min with the primary antibody raised against pAKT-Thr308 (dilution 1:200, clone C31E5E, catalog number 2965, Cell signaling Technology, Danvers, MA). After three washes in Tris-buffered saline, slides were incubated for 30 min with Dako Envision+ Dual Link at room temperature. Following three additional washes, slides were incubated with Dako chromogen substrate for 5 min and were counterstained with hemotoxylin for another 5 min. The intensity and extent of cytoplasmic and nuclear pAKT-Thr308 immunostaining was evaluated using a light microscope (magnification, x20). Cytoplasmic expression was quantified using a four-value intensity score (0, none; 1, weak; 2, moderate and 3, strong) and the percentage (0-100%) of the extent of reactivity). A final cytoplasmic expression score for pAKT-Thr308 was obtained by multiplying the intensity and reactivity extension values (range, 0-300). Nuclear expression was quantified using the percentage of extent of reactivity which gave rise to a nuclear expression score for pAKT-Thr308 (range, 0-100). Representative pAKT-Thr308 immunostaining (upper, strong; bottom, weak) is depicted in **Figure 7A**.

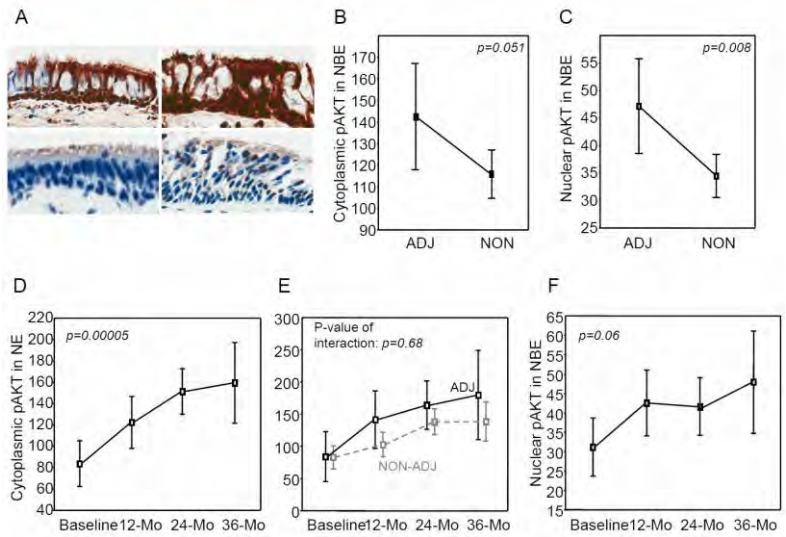

Figure 7. Immunohistochemical analysis of pAKT-Thr308 in airway biopsy specimens of the Vanguard cohort. **A**. Representative photomicrographs of pAKT-thr308 immunostaining (upper, strong; lower, weak). ANOVA analysis of cytoplasmic (**B**) and nuclear (**C**) pAKT-Thr308 immunohistochemical expression in adjacent (ADJ) and non-adjacent (NON) normal bronchial epithelia (NBE). ANOVA analysis of cytoplasmic (**D-E**) and nuclear (**F**) pAKT-Thr308 expression change by time in the Vanguard airway biopsies.

IHC expression of pAKT-Thr308 was compared across the examined airway biopsies based on site from the original location of the tumor as well as time the biopsy was performed following the baseline timepoint. Cytoplasmic pAKT-Thr308 expression exhibited a trend towards an increase in ADJ airways compared to non-adjacent and contralateral airways (NON) (**Figure 7B**). Nuclear pAKT-Thr308 IHC expression was significantly higher in adjacent compared to non-adjacent and contralateral airways (p=0.008 of ANOVA test) (**Figure 7C**). Interestingly, when we compared and contrasted pAKT-Thr308 expression based on time, ANOVA analysis revealed a significant up-regulation of the cytoplasmic expression of the phosphorylated protein with time in the airway biopsies (p=0.00005, **Figure 7D**). Moreover, the difference in the increase in cytoplasmic pAKT-Thr308 expression with time between adjacent and other airways was not significant following testing for the interaction of site and time variables by ANOVA (**Figure 7E**). Nuclear pAKT-Thr308 expression was also up-regulated with time in the analyzed airway biopsies, but was less significant when compared to cytoplasmic expression of the phosphorylated protein (**Figure 7F**). It is important to note that analysis of pAKT-thr308 was

performed in normal bronchial epithelia (NBE) only. Similar results were obtained when biopsies from main carinas were included in the analyses.

B. Gene expression analysis of bronchial epithelial samples obtained from lobectomy specimens from NSCLC patients (*Field Cancerization Study*):

To increase our understanding of the molecular basis of lung cancer pathogenesis, we analyzed the transcriptome profiles of cytological specimens (brushings) obtained at lobectomy or pneumonectomy procedures from 2-5 bronchioles with differential proximity from resected tumors and from resected lung tumors and normal parenchyma (**Figure 8**). Samples were obtained from patients undergoing lobectomy or pneumonectomy procedures (n=23) under an MD Anderson institutional review board (IRB)-approved protocol. A summary of the clinicopathological characteristics of the studied patient cases is depicted in **Table 1**. Tumor and normal parenchyma specimens for transcriptome profiling were collected using three different techniques per patient; brushing, snap-freezing in liquid nitrogen and preservation in RNA*later*. Brushings of tumor and normal parenchyma as well as bronchial brushings from airways were performed using Cytosoft brushes (CardinalHealth, Dublin, OH) and placed in Qiazol (Qiagen, Valencia, CA) and immediately placed in dry ice and stored in -80 °C until RNA isolation. Cytological quality control of the epithelial and malignant cell content of the collected histological tissue specimens was performed and is available. Total RNA of all samples was isolated using the miRNeasy Mini kit following homogenization of tissues and brushing collections. RNA quality was assessed by the 28S/18S ribosomal RNA ratio. 226 samples were eligible for microarray profiling based on RNA quality. Processed RNA samples were hybridized to Affymetrix GeneChip® Human Gene 1.0 ST arrays and were then scanned using the GeneChip® Scanner 3000 from Affymetrix (Santa Clara, CA) to yield raw image files that were subsequently converted to probe set data. Raw expression data was analyzed using

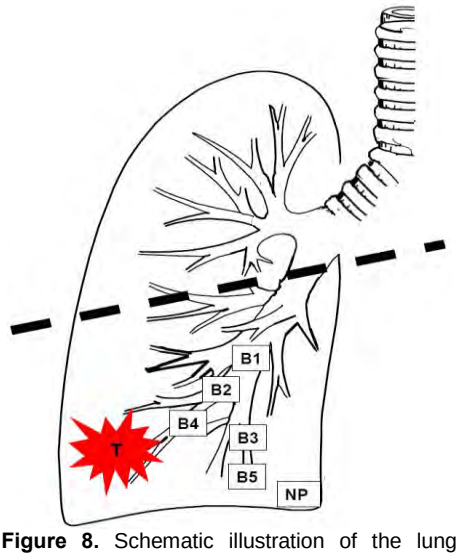

Figure 8. Schematic illustration of the lung airway for molecular mapping of field cancerization by high-throughput profiling. Tumor (T) and normal parenchyma (NP) samples were obtained by brushing, preservation in RNA*later* and snap-freezing in liquid nitrogen. Airway samples (B1 to B5) were obtained by brushing.

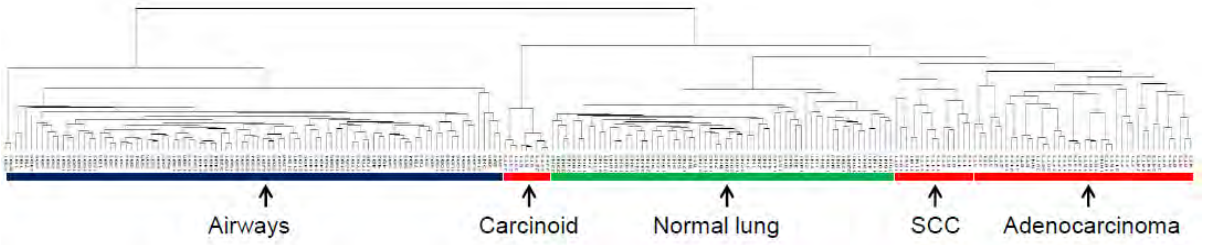

Figure 9. Hierarchical clustering analysis of the lung field cancerization samples. Unsupervised hierarchical clustering analysis of field cancerization samples (n=226) analyzed by microarray profiling. All genes present on the Human Gene 1.0 ST array (Affymetrix) were median centered and arrays clustered by average linkage. Airway (blue), normal lung (green) and tumor (red) major clusters are indicated by the color bars. Tumor sub-clusters (carcinoids, SCCs and adenocarcinomas) are indicated by name underneath the arrows.

the BRB-ArrayTools v.4.1.0 developed by Dr. Richard Simon and the BRB-ArrayTools Development Team and then normalized and log-transformed by the robust multi-array analysis (RMA) method using the R language environment.

We first analyzed the entire set of collected and microarray-profiled field cancerization samples (n=226) by unsupervised hierarchical clustering analysis. The entire set of genes present on the Human Gene 1.0 ST array (n=33,251) were median-centered and samples were clustered by average linkage. The unsupervised clustering revealed the presence of major clusters, as indicated by the colored horizontal bars (airways, blue; normal lung, green; tumors, red) based on histopathology of the specimen analyzed (**Figure 9**). Interestingly, the tumor cluster was further divided into sub-clusters based on the type of lung tumor profiled (carcinoids, SCCs and adenocarcinomas). In addition, all carcinoid tumors (n=3 cases, 9 samples) grouped into an independent sub-cluster and all NSCLCs were divided into two major sub-clusters with one harboring entirely lung SCC cases and the second lung cluster including all lung adenocarcinoma cases as well as a NSCLC (sarcomatoid) case.

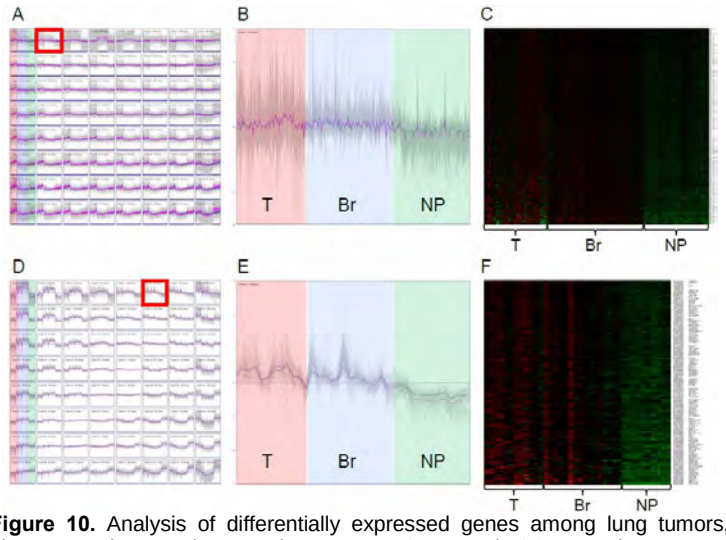

Figure 10. Analysis of differentially expressed genes among lung tumors, airways and normal parenchyma. ANOVA and SOM analyses were performed for all tumors (**A**, **B** and **C**) and SCC cases alone (**D**, **E** and **F**). SOM analysis was to generate 8 x 8 neuronal clusters (**A**, all tumors; **D**, SCCs). Representative clusters are highlighted in **A** and **D** and depicted in **B** and **E**. Median-centered expression levels of genes within the representative clusters are depicted in the heatmaps in **C** and **F**.

It is important to note that, in most cases, tumors and normal parenchyma samples collected by brushing, preservation in RNA*later* and snap-freezing in liquid nitrogen sub-clustered in groups of 3 by patient indicating that differences within profiles of tumors and normal parenchyma collected by the three methods (intra-group differences) are smaller than differences between profiles of tumors and normal parenchyma across patients (inter-group differences).

We then sought to identify genes differentially expressed among tumors, airways (bronchial brushings and normal parenchyma. We applied ANOVA using a p-value of 0.001 as a threshold for statistical significance. Significant genes identified by ANOVA were then queried using self-organizing map (SOM) analysis to identify clusters of genes displaying expression among the groups. SOM analysis was performed using Genesis software developed by Alexander Sturn and Rene Snajder (Graz University of Technology, Graz, Austria) using grids of 8 x 8 neuronal clusters (**Figures 10A and 10D**). SOM analysis gave rise to gene clusters with different variations of expression among tumors, airways and normal lung parenchyma (**Figure 10**). This analysis enabled us to highlight genes with an increased expression from normal parenchyma to airways and to tumors when all tumor cases (n=23) were analyzed (**Figures 10B and 10C**) or when only lung tumors, airways and normal lung parenchyma corresponding to SCC cases (n=5) were analyzed (**Figures 10E and 10F**). The representative cluster depicted in **Figure 10B** was found to be comprised of 362 genes that displayed increased expression in both lung tumors and airways compared to normal lung when all cases (n=23) were analyzed (**Figures 10B and 10C**). A representative cluster depicted in **Figure 10E** was found to be comprised of 140 genes that displayed increased expression in

lung SCCs and corresponding airways compared to matched normal lung (**Figures 10E and 10F**). It is noteworthy that the aforementioned analysis was successful in identifying genes with significant different variations of expression, when applied to sub-groups of the dataset, suggesting that such approaches may be useful in understanding the molecular profiles of the field cancerization phenomenon in relation to histology, e.g. lung adenocarcinoma compared to SCCs. Importantly, our data highlighted genes that are up-regulated in both lung airways and tumors compared to normal lung tissue and, therefore, may play important roles in lung cancer pathogenesis and serve as potential targets for chemoprevention.

Table 1. Clinico-pathological features of NSCLCs cases examined in the Field Cancerization Study

Covariate	Levels	N (%)	Covariate	Levels	N (%)
Gender	Female	11 (47.8%)	T	1	8 (34.8%)
	Male	12 (52.2%)		2	12 (52.2%)
Race	Asian	2 (8.7%)		3	0 (0.0%)
	African-American	4 (17.4%)		4	3 (13.0%)
	Caucasian	17 (73.9%)	N	0	20 (87.0%)
Tobacco history	No	8 (34.8%)		1	2 (8.7%)
	Yes	15 (65.2%)		2	1 (4.3%)
Smoking	Current	6 (26.1%)		3	0 (0.0%)
	Former	9 (39.1%)	M	0	23 (100.0%)
	Never	8 (34.8%)		1	0 (0.0%)
Histology	ADC	14 (63.6%)	Stage	I	17 (73.9%)
	SCC	5 (36.4%)		II	2 (8.7%)
	Carcinoid	3 (13.0%)		III	4 (17.4%)
	NSCLC NOS	1 (4.3%)		IV	0 (0.0%)

ADC, adenocarcinoma; NSCLC, non-small cell lung carcinoma; NOS, non-otherwise specified

The bioinformatic analysis of the aforementioned field cancerization pilot dataset (23 patient cases, 226 samples) is ongoing and is expected to be completed by January 2012 (Table 1). Field cancerization profiles will also be analyzed based on smoking status (e.g. lung adenocarcinoma smokers versus non-smokers). Field cancerization profiles will also be analyzed to potentially unravel genes displaying a site-dependent effect in relation to the original resected primary tumor. In addition, genes differentially expressed among lung tumors, airways and normal lung tissue will be functionally analyzed and topologically organized by pathways and gene-interaction network analyses. Moreover, field cancerization profiles of lung adenocarcinoma and SCC cases will be compared and contrasted in an attempt to identify genes and cell signaling pathways that may play important roles uniquely between the two major subtypes of NSCLC. Collection of lung tumor, normal parenchyma and airway samples is ongoing to profile a field cancerization set comprised of a larger number of patients.

Specific Aim 2: Integrate field cancerization profiles with available gene expression data of bronchial brushings isolated at various time points post-

surgery (DoD VITAL cohort), and correlate with tumor recurrence and metastasis.

Summary of Research Findings:

Dr. Wistuba's lab is not part of this Aim.

Specific Aim 3: Identify significantly modulated biological gene networks and pathways as well as differentially expressed network hub genes and associated miRNAs that might play key roles in pathogenesis of NSCLC.

Summary of Research Findings:

The studies on this Aim will be carried out in Years 3 and 4 of the grant.

Key Research Accomplishments:

- Identified that gene expression is modulated in a site- and a time-dependent manner in the bronchial epithelium of early stage lung cancer patients.
- Identified several pathways preferentially activated in the airway adjacent to tumors in patients with lung cancer, including those mediated by PI3K, NF-kB and ERK1/2.
- Completed the collection and field cancerization gene expression analysis of 23 patients (n=226 samples) with lung tumors using samples obtained from lobectomy specimens.

Conclusions:

During our first year of research, we demonstrated a localized field cancerization phenomenon on gene expression in the airway of patients with lung cancer, and we identified several pathways preferentially activated in the airway adjacent to tumors. We will continue to perform sample collections and data analysis of field cancerization specimens obtained from surgically resected lungs from patients with lung cancer to further examine the localized field cancerization phenomenon in the distal airway. In addition, our analysis will allow us to identify genes and cell signaling pathways that may play important roles uniquely between the two major subtypes of NSCLC, adenocarcinomas and squamous cell carcinomas.

Reportable outcomes:

Kadara H, Saintigny P, Fan Y, Chow CW, Chu ZM, Lang W, Behrens C, Gold K, Liu D, Lee JJ, Mao L, Kim ES, Hong WK, Wistuba II. Gene expression analysis of field of cancerization in early stage NSCLC patients towards development of biomarkers for personalized prevention. Proceedings of the 102nd Annual Meeting of the American Association for Cancer Research; 2011 Apr 2-6; Orlando, Florida. Philadelphia (PA): AACR; 2011. Abstract #3674.

Wistuba I, Kadara H, Kim ES, Hong WK. Molecular Pathology of Lung Cancer & Intermediate Markers of Carcinogenesis. 14th World Conference on Lung Cancer; 2011. Abstract #M19.